Auditory Displays and Assistive Technologies: the use of head movements by visually impaired individuals and their implementation in binaural interfaces

Thesis submitted in partial completion of the degree of

Doctor of Philosophy (PhD)

in

Human Computer Interaction

De Montfort University

By

Christopher Feakes

Leicester Media School

Faculty of Computing, Engineering and Media

De Montfort University

November 2018

Auditory Displays and Assistive Technologies: the use of head movements by visually impaired individuals and their implementation in binaural interfaces

Abstract

Visually impaired people rely upon audition for a variety of purposes, among these are the use of sound to identify the position of objects in their surrounding environment. This is limited not just to localising sound emitting objects, but also obstacles and environmental boundaries, thanks to their ability to extract information from reverberation and sound reflections- all of which can contribute to effective and safe navigation, as well as serving a function in certain assistive technologies thanks to the advent of binaural auditory virtual reality.

It is known that head movements in the presence of sound elicit changes in the acoustical signals which arrive at each ear, and these changes can improve common auditory localisation problems in headphone-based auditory virtual reality, such as front-to-back reversals. The goal of the work presented here is to investigate whether the visually impaired naturally engage head movement to facilitate auditory perception and to what extent it may be applicable to the design of virtual auditory assistive technology.

Three novel experiments are presented; a field study of head movement behaviour during navigation, a questionnaire assessing the self-reported use of head movement in auditory perception by visually impaired individuals (each comparing visually impaired and sighted participants) and an acoustical analysis of inter-aural differences and cross-correlations as a function of head angle and sound source distance.

It is found that visually impaired people self-report using head movement for auditory distance perception. This is supported by head movements observed during the field study, whilst the acoustical analysis showed that interaural correlations for sound sources within 5m of the listener were reduced as head angle or distance to sound source were increased, and that interaural differences and correlations in reflected sound were generally lower than that of direct sound. Subsequently, relevant guidelines for designers of assistive auditory virtual reality are proposed.

Foreword Regarding the Structure of this Document

This thesis is organised into four sections, each containing multiple chapters concerning a topic, or area of work. An outline of these sections and their content is listed below:

I Background:

- Introduction to the thesis
- Core Notions, including the fundamentals of acoustics, and psychoacoustics as required to read this thesis
- Literature Review, concerning the state of the art in human auditory localisation study
- Contextual Background, concerning the state of the art (and this work's place) within the field of virtual reality

II Field study:

- Data Analysis Toolkit, describing the design and implementation of a sensor package for use in field studies of human head movement
- Field Study, describing the design and results of a comparative study of human head movement during navigation for sighted and visually impaired individuals

III Distance Perception Studies:

- Self-report questionnaire Study of the Visually Impaired, describing the design, application, and results of a questionnaire regarding auditory perception-based navigation, and head movement, in the visually impaired
- Analysis of Binaural Cues in Reverberant Space, describing the design, implementation, and results of an acoustical study conducted to assess the impact of sound source distance and head movement upon auditory localisation cues

IV Conclusions:

- A discussion of the studies conducted, illustrating the scope of the data and results gathered, and their implications
- A concise guideline for the design of assistive auditory virtual reality, including some example systems to illustrate and justify the choices suggested by the guideline
- A discussion of contributions and possible future work arising from this thesis

Acknowledgements

My supervisory team: Lorenzo Picinali, Dylan Menzies, and Mark Scase. RNIB Connect, and Vista Leicestershire & Rutland for their assistance in recruiting test participants, Ken Clegg for his advice and teaching on electronics, and my family and friends for knowing better than to ask!

Table of Contents

Table of Figures

Abbreviations/Glossary

9DoF	Nine degrees of freedom
AIR	Approximated impulse response
ADC	Analogue to digital conversion
AHRS	Attitude heading reference system
BEMR	Bootstrapped estimate mean response
BMLD	Binaural masking level difference
BRIR	Body Related Impulse Response
ERP	Event-related potential
DTFT	Discrete-Time Fourier Transform
DRR	Direct to reverberant ratio
GI	Global interpolation
HAL	Human auditory localisation
HRIR	Head-related impulse response
HRTF	Head-related transfer function
IID	Inter-aural intensity Difference
ILD	Inter-aural level difference
IPD	Inter-aural phase difference
IR	Impulse response
ITD	Inter-aural time difference
KEMAR	Knowles Acoustic Manikin for Auditory Research
LI	Local interpolation
LSST	Logarithmic sine sweep technique
MAA	Minimum audible angle
MLS	Maximum length sequence
PD	Pure Data
RMS	Root-mean-square
RNIB	Royal National Institute for the Blind
RT_{60}	Reverberation time (duration)
SPL	Sound pressure level

STFT	Short-Time Fourier Transform
TSP	Time stretched pulse
VAD	Virtual auditory display
VI	Visually impaired
VR	Virtual reality

Part I: Background

1. Introduction

Much work in the fields of acoustics and psychoacoustics has focused upon examining human sound localisation- to what extent can we determine the source of a sound, and how do we achieve this [1]? As techniques for virtualising environments and sounds have become more viable the setting for many such psychoacoustical studies has moved into the virtual world, where researchers are able to extensively control the environment [2]. But, virtual reality has brought with it questions of user interaction and behaviour [3] [4] [5], amongst these are the questions of head movement- how should a virtual world respond when the user moves their head? Does head movement facilitate some aspect of audition that is particularly relevant in a virtual environment [5] [6]? Do people tend to use head movement to facilitate audition naturally? The latter question is scarcely approached in literature regarding the design of virtual reality systems, although the design of studies in this area implies an assumption that people do indeed naturally use head movement in this way [5] [6].

This question is particularly relevant to assistive auditory virtual reality systems, where questions over the applications and benefits or drawbacks of emulated head movement and head tracking technology are prevalent [7] [8] [6], as is the need for the simplest (and by extension most affordable [9]) functional system possible. According to the Royal National Institute for the Blind there are an estimated two million people living with some level of sight loss in the United Kingdom alone [10]. For the visually impaired, issues of accessibility of both environment and technology have been cited as barriers to work, and social life [11] [12] [13] [14].

The use of head movement is explored here in terms of whether or not it is naturally used by visually impaired people when navigating environments and consequently, testing the apparent assumption that it is, and investigating how it may be implemented into virtual reality systems (with particular attention paid to virtual environments intended to assist the formation of cognitive maps, or to create spatial auditory representations of traditionally visual information) to facilitate the use of such systems.

1.1 Auditory Virtual Reality

Computer based Virtual Reality (VR) technology has existed since at least 1968, when Ivan Sutherland introduced the first head mounted computer display intended to allow users to occupy the same space as virtual objects rendered with simple computer graphics [15]. By the 1990s virtual reality systems were being employed as training aids in fields where immersive computer simulations could be used to prepare individuals for potentially hazardous activities, such as military operations [16].

The evolution of immersive audio in virtual reality technology, facilitated by steady increases in available computational power, has led to the development of 'audio-centric' virtual applications and equipment. So-called Virtual Auditory Displays (VADs) have a range of applications and research may be found towards virtual auditory displays operating in conjunction with computer system to improve or assess accessibility for the visually impaired [17] [18] [19].

1.1.1 Binaural Audio Rendering

To immerse the user in a simulated auditory scene which mimics reality, it is necessary to create a sense that sounds are coming from a variety of locations around the user. In a traditional stereo speaker system, for example, sounds may be played from either one, or both of a pair of speakers. If the sound is played from both speakers with equal intensity ('loudness') then the sound should appear to originate from a point directly between the speakers. If the sound is presented with greater intensity at one speaker than the other, then the sound will seem to originate from a point closer to the 'louder' speaker. By adding more speakers around the listener, more potential locations for the origin of a sound are added (this is the principle behind 5.1 and 7.1 surround sound technologies) [9].

This method has limitations: it is very difficult to simulate distance accurately as all sounds originate from the exact distance of the speaker that produces them [9], and numerous speakers are required to accurately create a 3D sound scape (with the ability to simulate vertically as well as horizontally located sounds [9] [20]).

Binaural audio rendering offers a possible solution to both problems. By recreating the psychoacoustical cues which are used by the human brain to determine the location from which a sound has originated (discussed further in *chapter 2*), binaural rendering has the potential to allow sounds to be placed into a fully realised 3D 'scene' around the listener [21]. As the cues are encoded into individual sounds it is only necessary to have headphones reproducing the signal; binaural rendering is most commonly presented via a pair of headphones. For this reason, perhaps, binaural audio has become of great interest to researchers and designers of virtual reality and the relative cheapness and simplicity of the presentation hardware compared to some multi-speaker methods makes it viable for consumer use.

1.1.2 Assistive Binaural Virtual Reality

Assistive binaural technology has been explored or suggested for many applications; systems aimed at allowing a user to form and learn mental maps of a particular environment (such as an office or public building), sonified computer desktops and software interfaces to improve accessibility to technology, and simulations for experiencing and practicing dangerous scenarios such as road crossings. In addition to these applications, augmented reality systems using spatial audio for sonified waypoints in real world navigation have also been explored (these aspects are discussed further in *chapter 4*).

That none of these technologies have developed in to what could be considered a standard system, nor have they become widely adopted in a commercial sense, is an indication that there are still issues surrounding their implementation. Throughout the literature of this thesis several current questions regarding the most effective way to implement binaural audio in virtual reality are presented:

- How may head movement affect auditory perception?
- How can head movement be effectively employed in a virtual reality system?
- What role do environmental acoustics play in auditory perception?
- How can common problems in binaurally rendered audio be overcome?

There are relationships between some of these problems, which will become clear throughout the discussion given in this dissertation, and in some respects the answers to some of these questions have perhaps already been approached. A synergy between technological possibilities and psychoacoustics can be achieved, in which a binaural virtual audio system fulfils the perceptual needs of the user sufficiently to create a stable, auditory spatial illusion by incorporating the necessary technical facets to produce viable psychoacoustical cues without implementing excessive technical features which have inconsequential perceptual impact.

1.2 General Methodologies

The primary research presented in this dissertation was conducted in three stages. The first, a field study, was conducted using a novel head tracking system which was attached to the heads of visually impaired (VI) and sighted participants, who were then instructed to navigate a route through an urban environment whilst the system recorded their head movement behaviour. These measurements were assessed descriptively, by comparing head turn frequency at different stages of the route, and in the presence of different environmental factors. It was hoped that a significant difference in behaviour between visually impaired and sighted individuals might be found at this stage, such that a statement towards the use of head movement in auditory perception could be drawn out.

Unfortunately, no such difference was clearly and consistently observed, possibly due to the small size of the study group (discussed further in this chapter) or simply because no other significant difference exists. Some qualitative differences in the use of head movement between sighted and visually impaired participants were noted, particularly in relation the extent of yaw rotations at road crossings. Some indication of head movement use was observed in both groups, as well as a difference in head turn velocity between sighted and visually impaired participants under certain conditions. Post-test debriefing of two visually impaired participants further indicated that head movement may facilitate auditory perception of distance, and perhaps sound source movement.

The second study explored the possible use of head movement via a self-report questionnaire for the sighted and visually impaired (with differentiation between those affected by late and early onset of sight-loss). The questionnaire used Likert scales to test respondent's awareness of the use of head movement for auditory perception, placing focus upon distance, movement perception and localisation confidence, but also asking for further information relevant to the study topic in the form of open questioning. A principal component factor analysis and statistical resampling (the bootstrap method discussed later in *chapter 7*) was used to infer the relevance and power of the results to the visually impaired population, where it was found that head movement was reportedly used to facilitate auditory perception.

The final study was an acoustical experiment measuring potential spectral and interaural acoustical cues for distance with a KEMAR[1] mannequin in reverberant space, with a sound source placed at a range of distances from 1m to 10m, and a variety of head angles tested. The sound source was a logarithmic sinusoidal sweep presented form a geodesic, omni-directional speaker. Results were compared for ILDs as well as direct to reverberant energy ratios (as these were considered important for distance perception), and spectrograms were used to further explore spectral changes as a function of head angle and sound source distance. This experiment was somewhat an expansion upon similar studies in literature, where the impact of environmental acoustics upon HRTFs has emerged as an area of interest. In the context of the work presented here it was of interest to measure the possible changes in localisation cue in the presence of reverberation/sound reflection under the conditions of both head movement and varying sound source distance.

1.2.1 On Small Sample Sizes

Two studies presented in this work; the field study of head movement behaviour (in *chapter 6*) and the self-report questionnaire study of the visually impaired (in *chapter 7*) contained small sample groups: the field study was conducted with a total of eight participants and the questionnaire with a total of 26. As is discussed further in the respective studies' chapter, the small sample sizes were in part due to ethical

[1] An anthropomorphic mannequin commonly used in psychoacoustical studies

considerations placing limitations upon participant eligibility, and partly due to the relative sparsity of the visually impaired population.

Small sample sizes such as these are not uncommon in research dealing with of developing virtual reality systems, auditory perception, or cognitive mapping, particularly where work concerns the visually impaired. Key studies in these areas (discussed in *chapter 4*), such as Begault [22], Wenzel [3], and Katz & Picinali [8] have been conducted with as few as five participants, in such cases statistical techniques comparing variance (such as analysis of variance) between or within groups is the most common statistical method employed on non-resampled data, and is that which was used in analysing the field study presented in this work.

1.3 Contributions

Previous research and literature on head movement and auditory perception (presented throughout *chapter 3 & 4*) tends to presuppose that head movement plays a role in some aspect of audition and, by design, either guides human participants towards the use of it, or to attempt to study audition in the absence of head movement- placing physical restrictions or limitations upon participants' movement via the apparatus employed during study. New studies into the use of head movement in auditory perception by the visually impaired have been presented here, in which the use of head movement (or a lack thereof) is of primary concern. This research has contributed some way towards answering the question of whether visually impaired persons naturally use head movement to facilitate audition, and what function they self-report that head movement serves- this is perhaps the first time such a question has been posed, or an answer been approached.

This work was guided by the general hypothesis that head movement would be observed in natural use by visually impaired individuals who had not otherwise been guided to use it. Such movement was certainly observed during navigation. Further, neither the self-reported nor observed, use of head movement varied significantly between sighted and visually impaired individuals outside of greater head yaw velocities measured in the visually impaired group. In both measured and self-reported conditions, it was the qualitative comparison which yielded some differences between the groups: the visually impaired self-reported that head movement was helpful in facilitating distance estimation, and reflected more upon their safety concerns when navigating (such as quiet vehicles). More generally the studies conducted and presented here indicated that the visually impaired *do* use head movement, that they self-report that it *is* beneficial to auditory localisation, but that there is little difference in these behaviours between sighted individuals and those with early or late-onset sight loss.

A guideline based upon the primary and secondary research of this text has been created with the goal of aiding designers of assistive binaural auditory virtual reality to reduce the complexity and cost of systems by considering what is functionally necessary for the end user and the intended functions of the system. This includes guidance towards the use of either head tracked, manually controlled, or emulated head movements. As current literature has already demonstrated that head movement can be used to ameliorate known deficiencies in binaurally rendered audio it is suggested that one of these forms of head movement be considered in any assistive virtual auditory environments which use it.

2. Fundamental Concepts

This chapter deals with some background necessary for reading the remainder of the dissertation. It is intended for those who are less familiar with concepts in human auditory localisation, and some phenomena (namely the cone of confusion and precedence effect) which are relevant to the primary and secondary research presented here. It is by no means a literature review but rather an effort to define in brief but certain terms that information which is most relevant to the topics upon which this thesis is founded.

2.1 Spatial Coordinate System

To discuss auditory localisation and acoustic events around a human listener, it is first necessary to define a common spatial coordinate system within which these phenomena can be described. This coordinate system is spherical and head related; meaning that it remains constant to the listener's head, and the head will always be located at the origin[2] of the system.

Three distinct planes divide the coordinate sphere:

- The horizontal plane: placed at the superior margins of the ear canals and the inferior margins of the ocular cavities.
- The median plane: placed at 90° to the horizontal plane, it aligns to the axis of symmetry of the head.
- The frontal plane: placed at 90° to both the horizontal and median planes, it intersects with the superior margins of the ear canals.

These three planes are coincident with the 3 degrees of freedom available to the human head; that it is possible to yaw (rotate) the head consistent to the horizontal plane, pitch it forwards or backwards in accordance with the orientation of the median plane, or roll it left to right in accordance with that of the frontal plane, or perform a complex movement which is a compound of these three.

---

[2] The centre of the spherical coordinate system

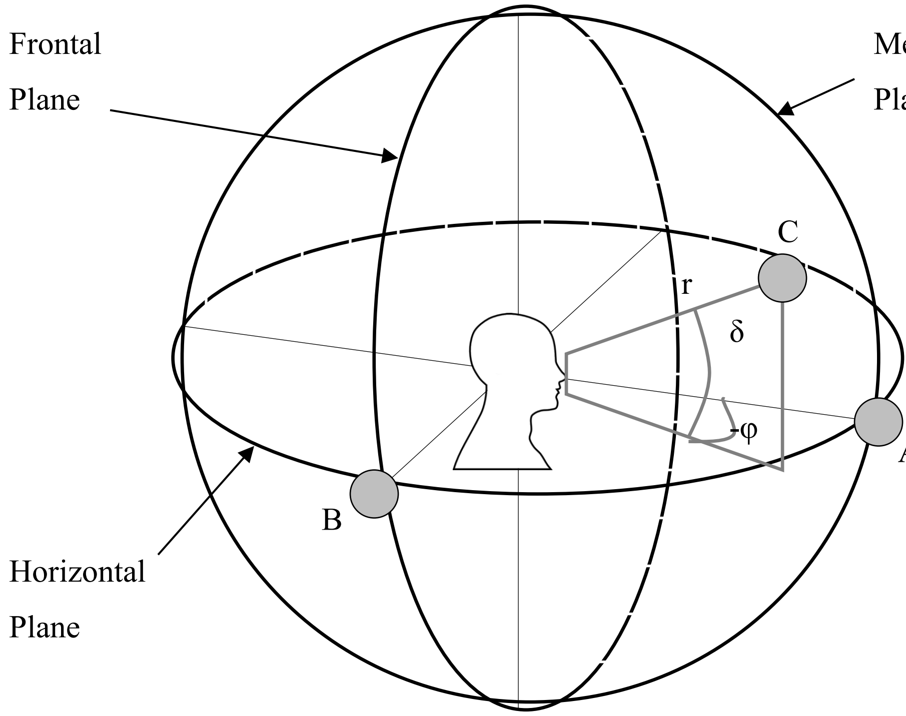


Figure 2.a An illustration of the coordinate system, with listener's head at the origin [213, 291]

A point within the spherical system can be defined by the Azimuth (φ, angle on the horizontal plane, proceeding clockwise from point A in *figure 2.a*), elevation (δ, angle on the median or frontal plane, proceeding upwards), and distance (r, the distance from the origin to the point in question). The coordinates of points A, B, and C in *figure 2.a* would be given as:

A: $\varphi = 0°$, $\delta = 0°$, $r = 1$ (in the case of a unit sphere)
B: $\varphi = 95°$, $\delta = 0°$, $r = 1$(in the case of a unit sphere)
C: $\varphi = 15°$, $\delta = 30°$, $r = 1$(in the case of a unit sphere)

2.2 Human Auditory Localisation

Human auditory localisation (HAL) is achieved via the perception of two major sets of psychoacoustic cues, interaural time differences (ITD)[3], and interaural level differences (ILD)[4]. This so-called 'duplex' theory of sound localisation was first quantified by Rayleigh [23]. As this chapter will outline these are binaural cues and are reliant upon the arrival of a sound at both ears. A certain amount of localisation data is also available from monaural ([perceived via] one ear) cues such as spectral shaping caused by physiological features including the pinna, which facilitate the estimation of sound source elevation [24] [25]. As such cues are purely acoustic, relating to the external physiology of the hearing system and head/body, they are somewhat unique to the individual. The combinations of these localisation cues, along with other minor cues that are available in some circumstances (such as the Doppler effect [26]) create a non-uniform auditory perceptual spatial resolution around the listener. This allows listeners to discern the origin of a sound, and information regarding its movement, via the auditory channel with varying levels of accuracy; depending upon the location of the sound source relative to the listener's position/orientation, and factors such as movement speed and direction [27] [28]. It should be noted that when describing localisation this chapter refers to the estimation of angle/direction of a sound source, the perception of distance (or range, from the listener's perspective) is dealt with as a special case in chapter *3.4.* Each sub-section of this chapter deals with the perception of stationary sound sources, except for *chapter 3.5*, which relates specifically to the perception of sound source movement.

2.2.1 Binaural Localisation Cues: Interaural Time and Level Differences

The IID and ILD cues are the means by which the angular position of a sound source, relative to the listener, can be determined [28]. There are certain limitations, discussed throughout this chapter, as to the frequencies and angle of arrival of sound, at which these cues are effective.

---

[3] Also referred to as interaural phase difference (IPD) in some literature.
[4] Also referred to as interaural intensity difference (IID) in some literature.

2.2.2 Interaural Time Differences

The interaural time difference represents the difference in arrival time of a sound between one ear and the other. This difference occurs because of the physical separation of the ear canals and is created by the differences in distance a sound wave originating from a given point must travel to reach each of the ears in turn. It has been identified that the ITD is primarily determined by the phase difference that occurs when the same sound wave reaches the ears at different times [29], shown in the formula:

$$ITD = \frac{r(\emptyset + \sin(\emptyset))}{c}$$

There are limitations to the viability of ITDs as localisation cues. The first relates to the maximum frequency at which this cue is effective. This limitation arises when the difference in time of arrival is equal to or greater than a 180° phase shift in the sound wave, due to the diameter of the head being equal to or greater than one half wavelength, causing uncertainty as to whether the phase shift is positive or negative, and therefore an ambiguity as to which ear the sound arrived at first. Howard and Angus [29] propose the following equation for calculating the maximum frequency at which effective ITD cues arise:

$$fmax(\emptyset) = \frac{1}{2 \times 0.09m \times (\emptyset + \sin(\emptyset))}$$

It has been noted that humans can detect ITD in high frequency sounds by means of ‘envelope delay’ that is the disparity between signal amplitudes/signal onset at each ear [30, 31] although this cue is considered perceptually less significant than the detection of phase differences.

The second limitation is that the ITD only exists as an accurate localisation cue when the position of the sound source is not located directly upon the median plane, otherwise there will be no difference in arrival time between the ears, the phase of the sound arriving directly at each ear will be identical and ambiguity in the location of the sound upon the median plane may arise unless another cue is present.

Head movement could be used to dynamically change ITDs by increasing or decreasing the angular extent between the median plane and the sound source. Indeed, it has been observed that in pure tone localisation experiments where participants were free to move their heads that sighted participants often turned their heads slightly to help resolve the location of sounds placed at either 0º or 180 º azimuth [32], of course such movement would not only affect the interaural time differences, but also the interaural level differences.

2.2.3 Interaural Level Differences

The interaural level difference arises because of the shadowing and scattering effect of the listener's physique [29] where the presence of the listeners' head, body and pinna create an acoustic impedance which leads to differences in the relative level (intensity) of sound at each ear. The pinnae are of special importance when differentiating between sound sources located at 0°or 180° azimuth as they provide the primary difference in acoustic impedance and reflection of sounds arriving from directly in front of, or behind the listener. In this special case level differences created by the pinnae are considered monaural as an individual pinna can perform this reflection and impedance of sound, creating a sufficient cue for a listener to differentiate between front and rear located sources with a single ear [33].

Based upon the principal that an object will not significantly impede the propagation of a sound wave unless it is approximately greater than or equal in size to 66% of the sound's wavelength, Howard and Angus [29] offer the following equation to determine the lower frequency limit at which the human head (with a given size of 18cm from ear to ear) will no longer impose ILD cues:

$$fmin\left(\emptyset = \frac{\pi}{2}\right)\frac{1}{3}\left(\frac{c}{d}\right) = \frac{1}{3} \times \left(\frac{344\ ms^{-1}}{0.18\ m}\right) = 637Hz$$

The minimum threshold of sensitivity to ILD is 1dB [34] with ILDs of up to 35dB measured at 10kHz [35].

2.2.4 The Cone of Confusion

The cone of confusion [36] describes a series of spatial coordinates originating at the listeners' head with its central axes at ±90° azimuth and 0° elevation, where sound sources would produce identical ITD and ILDs at multiple positions on the cone. This means that were it not for the pinna, it would be impossible for the listener to accurately locate the position of the sound. Reversals in perceived location would occur, such as front-to-back reversals, and the listener may only be able determine whether the sound lay to the left or right of the median plane.

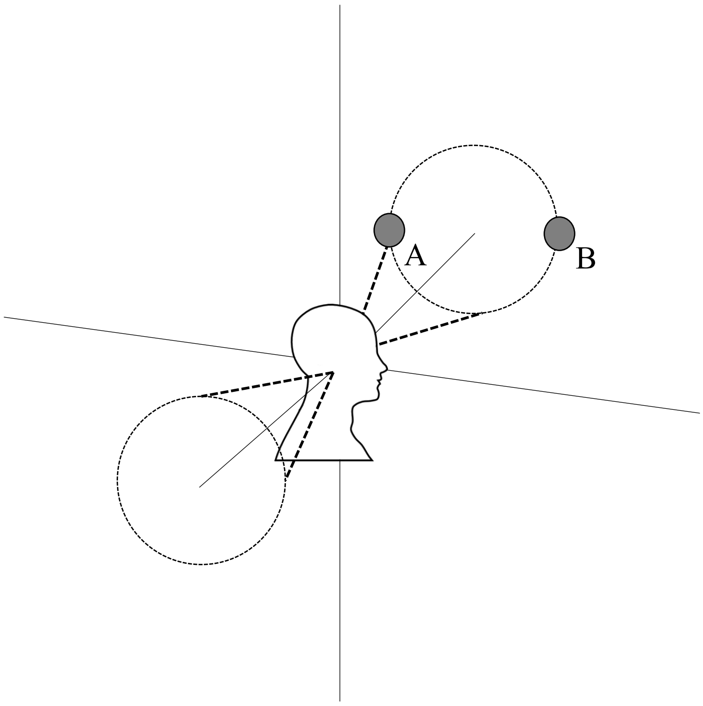


Figure 2.b The cone(s) of confusion shown relative to a listener's head. Points A and B are equidistant from both ears and so have identical ITD and ILD values, despite being on opposing sides of the frontal plane.

The monaural cues provided by the filtering effect of the pinnae can mitigate front-to-back reversals providing the sound in question is of appropriate spectral content [37].

The cone of confusion is another elegant example of a reason why head movement might be used to assist auditory localisation; localisation ambiguities for those sounds which fall upon it could be resolved with head movement, which would move the cone of confusion relative to the direction of the sound source.

2.2.5 The Precedence Effect[5]

When listening in a bounded acoustic environment such as a room or open space with acoustically reflective objects nearby, a challenge to accurate sound source localisation arises in the form of reverberation or sound reflection. As the diagram below illustrates, sound reflections arriving at the listening position could potentially be confused with additional sound sources around the listener:

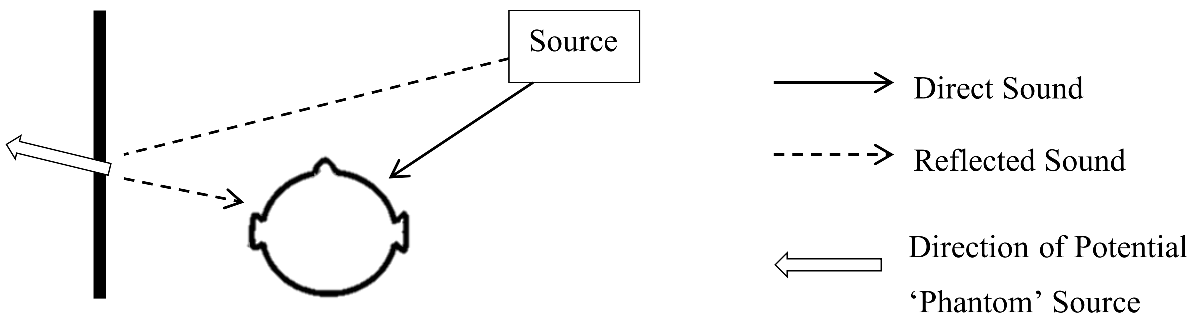


Figure 2.c A diagram illustrating the potential confusion of a sound source's location arising from reflected sound.

The precedence effect is the phenomenon (or compound phenomena) in which identical, or near identical, complex signals arriving from apparently different locations are perceived as a single sound emanating from the source. The source is determined by the localisation of the first wave front to arrive at the listening position. This phenomenon typically occurs when there is a difference in arrival time of less than 50-100 milliseconds, depending upon the nature and complexity of the sound in question [38].

---

[5] Sometimes referred to as the Haas Effect, Law of the First Wavefront, or Auditory Suppression.

In this way, the presence of a phantom sound source location associated with reflections of sound can be suppressed to prevent sound localisation in reverberant environments from being confounded.

2.2.6 Summary of Background Notions

The information presented here offers a basic background to the terminology which is presented throughout this dissertation. The coordinate system described is useful in the reading of both the literature and primary research described, whilst the description of auditory localisation and the phenomena of the cone of confusion and precedence effect are useful in the reading of the literature review.

In this fundamental description of the major features of human auditory localisation, possible use cases for head movement emerge. These will be further elaborated upon throughout *chapter 3.*

3. The State of the Art in Human Sound Localisation

Rayleigh's Duplex Theory of Sound Localisation [23] was perhaps the first treatise on audition which considered that sound source localisation was possible using the auditory channel alone, with prior approaches to studying audition tending to consider localisation impossible without use of the visual channel [39]. Whilst efforts to thoroughly quantify human auditory localisation continue, its existence cannot be denied. This chapter contains an overview of the continuing research carried out within auditory localisation, the psychoacoustic cues which underlie it, and its application to auditory virtual reality.

Due to the nature of the original primary research carried out in support of this thesis, this chapter includes a review of studies relating to auditory localisation (presented in more specific detail than that given in *chapter 1*), which are considered relevant to the implementation of both virtual auditory reality, and an understanding of the potential roles, and use of head movement in auditory localisation and scene evaluation- although, there is little or no direct study of the latter topic present in current literature. As this chapter illustrates, auditory spatial perception is bound by inequalities, the understanding of which may offer insight into the role head movement plays in sound source localisation.

Since sound source direction and distance, although each components of sound source localisation, are discussed separately in this chapter the sub-sections dealing with angular localisation are presented as dealing with "localisation".

3.1 Localisation: Localising Individual, Stationary Sound Sources

Locating stationary sound sources was amongst the most prevalent topics of early study in human sound localisation, perhaps owing to its relative simplicity from an experimental point of view; requiring no apparatus capable of moving active sound sources in a controlled fashion.

As detailed throughout *chapter 2*, there are several phenomena in auditory perception which form the cues necessary for an individual to determine the location of a given sound source. As the work summarised here demonstrates, the perception of even a single sound source's location is a complex issue, the accuracy with which sounds can be located depends upon several factors including sound frequency/spectral complexity, sound duration, and the angle of source location relative to the azimuth. This complex and uneven spatial resolution may offer further insight into *why* head movement could aid auditory localisation.

3.1.1 Horizontal Plane Localisation

The localisation of stationary sound sources in the horizontal plane, when the sound's spectral content allows (as described in *chapters 2.2.2* and *2.2.3*), is primarily achieved via the interaction of the ILD and ITD [23, 28]. The resolution of horizontal plane localisation has been the subject of numerous studies conducted both in the real world, virtual reality and mixed reality conditions[6].

In his 1958 study of auditory spatial resolution, Mills [40] aimed to determine the minimum audible angle (MAA) difference using pure tones ranging from 250Hz to 10,000Hz, played to test participants via a speaker fixed to the end of swinging boom arm

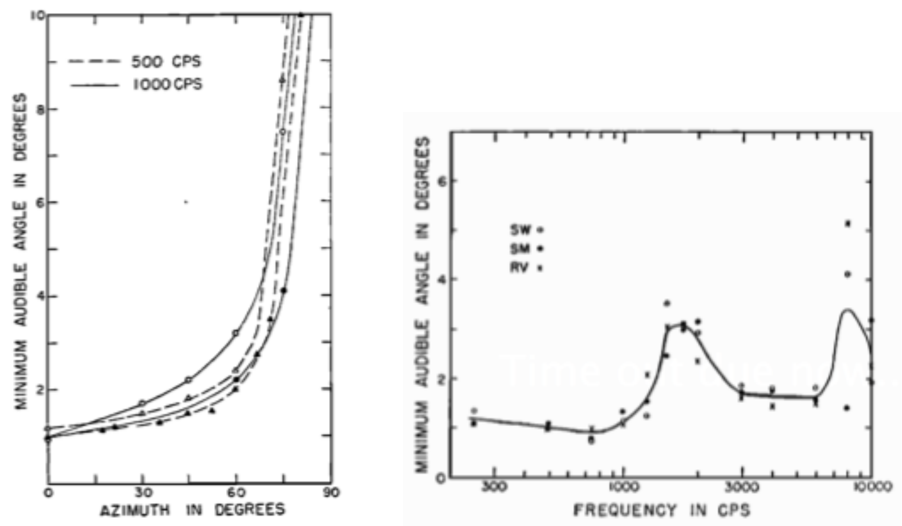


Figure 3.a LEFT: A graph showing the MAA as a function of the direction to sound source in the horizontal plane. RIGHT: A graph showing the MAA as a function of frequency (cycles per second) [28].

[6] Using augmented reality systems, or stimuli recorded in real environments and played back to test participants in a later study phase

in an anechoic[7] chamber. By playing sounds sequentially at slightly varying locations and asking participants to report the perceived direction of the sound source via a handheld rotary dial, Mills was able to make a number of observations: the minimum audible angle varied notably, with the smallest detectable difference of 1º in front of the listener with performance worsening as the sound source was moved lateral and rear positions relative to the azimuth. The minimum audible angle also varied depending upon the frequency of sound, with particularly high performance occurring at frequencies around ≈750Hz and again at ≈ 3,500Hz.

Such minimum audible angle studies remain a popular method of testing the auditory spatial resolution of humans, and have been translated into virtual studies of sound localisation. However, few of these studies have replicated Mills' original minimum resolution of 1º. This is perhaps due to his use of pure sine waves as stimulus, a practice which has somewhat given way to the use of more complex stimuli which are arguably more congruent with the complex natural sounds that humans encounter outside of experimental settings. More likely, however, is that it is a product of the virtual setting in which many studies are now conducted- It is known that there are a number of issues affecting binaural audio reproduction, often relating to the fact that binaural rendering is based upon measurements of human physiology which do not accurately match that of the test participant or user, meaning that from their perspective the psychoacoustical localisation cues in binaural audio do not correctly match their own (these issues are discussed further in *chapter 4*). Indeed studies which *have* replicated similarly low minimum audible angles to those of Mills have tended to be conducted outside of virtual reality [41] [42] [43].

In semi-virtual studies published in 2003, Grantham *et al* [44] employed the use of a dummy head[8] with microphones fixed inside the ear canals to record low, high and band-

---

[7] A room designed to prevent the audible reflection of sounds, achieved by the precise acoustic treatment of all potentially sound reflecting surfaces within.
[8] The KEMAR dummy head and torso system, which is used in audiological research to simulate the effect of human physiology upon sound, and to create ILD and ITD, cues in recorded sounds.

pass filtered noise within ranges from 70Hz to 15,000Hz. The stimuli were then presented to participants via headphones. The study tested the spatial resolution in horizontal, vertical and diagonal planes, and used an adaptive procedure whereby participants were asked to indicate whether two sequentially presented sounds appeared to be separated in angle, and whether the second sound was above, below, left or right of the original. By increasing the distance between stimuli until spatial separation was detected, and then reducing the angular separation until the sounds were perceived as emitting from the same place[9], thresholds of spatial resolution were determined. In the horizontal plane, measurements indicated that spatial resolution was independent of spectral content, with mean minimum threshold measurements of 1.6° recorded for wideband and high-pass stimuli, and 1.5° for low-pass stimuli. Although noticeably poorer than results obtained from real world studies, these were far from the worst obtained in virtual conditions, with some studies finding minimum audible angles from 4.8º [45] to 11° [17] in the horizontal plane.

A new dimension in the study of human audition would emerge, which would become particularly relevant in the context of current attempts to create assistive virtual reality. In a comparative study of auditory localisation involving congenitally blind and blindfolded, sighted individuals Roder *et al* [46] found auditory spatial perception in both groups to be similar in tasks involving discriminating between locations in front of the listener however, when asked to discriminate between sounds occurring in the 'peripheral' auditory space (those areas directly to the left or right of the participant), individuals from the blind group demonstrated improved performance compared to the sighted group. Event-related potential (ERP) neural imaging conducted during the tests was noted to indicate that these differences arose because of reorganised neural substrates[10] related to auditory attentive selection. Such studies comparing individuals with differing levels of visual acuity have grown in number as the issue of comparing the auditory localisation capabilities of blind and sighted individuals has grown, and a review

[9] A form of adaptive testing commonly used to determine thresholds in auditory perception.

[10] The brain tissue that underlies specific behaviors or psychological states

of studies specific to testing the comparative localisation abilities of visually impaired and sighted individuals can be found in *chapter 3.4*.

The inequality in sound localisation in the horizontal plane may provide cause for the use of head movement in facilitating localisation; placing sound sources of interest closer to 0° azimuth by moving the head should aid more accurate localisation whilst dynamic changes in spectral cues induced by the pinna may also help to resolve sounds which fall into the cone of confusion [47]. However, it has been noted that the visually impaired may respond to bilateral sound presentation on the horizontal plane by using it for postural control [48]- reducing head and body sway. This effect is not observed if sound is placed upon the median plane [14].

3.1.2 Median Plane Localisation

The localisation of sounds on the median plane, intersecting the horizontal plane exactly at 0° or 180°, presented a challenge to the Duplex theory- due to the symmetrical physiology of the human head there would be no potential for differences in level or time of arrival between the ears. Localisation in this plane was therefore assumed to occur without the use of inter-aural cues and was thought to be primarily the result of spectral filtering from the acoustic impedance offered by the pinna [49, 50]. Later studies revealed that, due to asymmetry of the pinna, ILDs *could* be measured in sources located vertically about the median plane. In fact, these differences were also prone to change with elevation of the sound source and could be as large as 10dB in certain frequency ranges [51, 52].

In a study of localisation cues in the median plane, Morimoto and Nomachi [53] found that participants could localise sounds in the median plane using spectral cues alone but were unable to localise sounds accurately using interaural differences caused by the pinna alone. When considered together however, the combination of interaural and monaural spectral cues yielded the most accurate localisation results.

In the vertical component of their 2003 spatial resolution tests Grantham *et al* [54] recorded the lowest spatial resolution in elevations on the median plane. Mean minimum thresholds of 6.5° for broadband noise, 11.6° for high-pass filtered noise, and 19° for low-pass filtered noise were recorded. However, [43] reported a mean minimum localisation threshold of 3.65° in response to stimuli consisting of a train of 50ms broadband clicks. It is important to note that, as stated previously, the pinna play a crucial role in vertical localisation. The use of a dummy head by Grantham *et al* to record stimulus may have affected the reported thresholds, since the pinna of the dummy are unlikely to precisely recreate the complex filtering effect of each test participant's own pinna. This is an issue critical to the implementation of binaural virtual reality (issues surrounding the matching of human physiology to its measured representations is discussed further in *chapter 4*).

Stimulus duration is another confounding factor here, which may account for differences in measured localisation accuracy. Differences in both subjective performance, and spectral psychoacoustical cues have been measured for stimulus of 50ms or less, when compared to longer stimuli which are similar in all other regards than duration [55] where mean localisation errors for stimuli of .5ms were as high as 70.9°, and for stimuli of .25ms the highest angular errors constituted almost total front-to-back reversal (178.3°). That [43] did not report such high angular errors is understandable, as theirs was a test of minimum audible angle movement thresholds, however in terms of absolute angular judgement, stimuli of the durations they presented could confound localisation in cases where spectral cues play a key role. This too presents a possible case for the use of head movement, rolling the position of the head to the extent that it introduces interaural differences (such a roll may in fact only require an angular extent of 10° to improve spatial resolution [43]) would eliminate the listener's reliance upon spectral cues. In reality such behaviour would depend upon the presence of such short duration sounds, and upon the listener's ability to react to them in a time frame sufficient for localisation to take place.

3.2 Localisation: Localising Multiple Sound Sources

Although a useful picture of auditory spatial resolution can be obtained through measurements using single, or multiple concurrent sounds, it is common for individuals to encounter complex sound-fields containing more than one sound source active at the same time. The study of localisation of multiple simultaneous sources, or localisation of single sources with the presence of some interfering sound or signal is therefore certainly of interest. This chapter will deal with literature regarding the localisation of both multiple sound sources simultaneously, and single sound sources in the presence of some other interfering sound source.

3.2.1 Localising Sounds Simultaneously

In his 1984 study of concurrent minimum audible angles, Perrot [56] employed the use of pure tones played simultaneously from two separate locations about the horizontal plane. Presented via speakers on swinging boom arms, the tones differed slightly in frequency, with the ‘target’ frequency fixed at 500Hz and secondary frequencies of either 515Hz, 543Hz, 572Hz, or 601Hz. Participants were asked to determine the direction of the target frequency relative to the secondary frequency (to the left, right, no difference).

Perrot noted that the accuracy with which participants reported the location of the target frequency was linked to both the difference between the target and secondary frequency, and the distance separating the two sounds. Performance improved as the spatial separation of the sounds increased, and as the difference in frequency increased. It was also noted that as the pair of speakers were swung to the lateral extremes of the horizontal plane, performance decreased slightly, replicating the results of [40]. In a virtual environment study of concurrent minimum audible angles, Best [57] presented listeners with binaurally filtered broadband noise. During these experiments the filters used were individualised to match the effects of each participant’s own outer ear. After repeating tests with components of the ITD removed (either differences in onset time, or differences in signal phase, or both), Best concluded that of the inter-aural cues ITD was the most significant in determining whether identical, broadband sound sources were spatially separated. Low frequency phase differences were found to be particularly useful when discriminating between sources located to the lateral extremes of the horizontal plane.

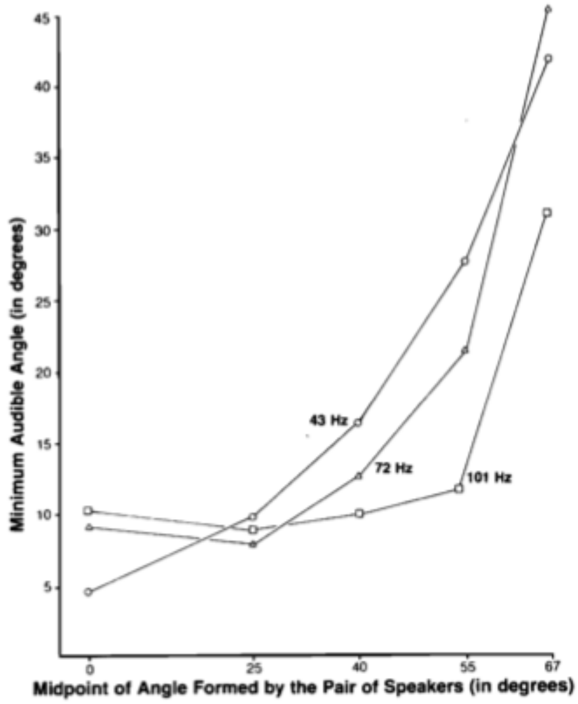


Figure 3.b A graph showing the minimum audible angle of concurrent sounds as a function of source position, separation, and frequency difference [291].

Moving beyond the study of two simultaneous sounds, Santala and Pulkki [58] conducted experiments involving an array of 13 stationary speakers, arranged along the frontal arc of the horizontal plane. Combinations of some, or all, speakers were used to simultaneously present band-pass filtered white noise, unfiltered white noise, or brown[11] noise, in different locations about the user. The study concluded that, in general, participants were unable to accurately identify the distribution of sound sources when more than three speakers were active and producing the same stimulus type. Improved performance in identifying speaker distributions was measured when the frequency range of stimulus was increased, with the most notable improvements achieved for stimulus in the ILD range. The study of multiple, simultaneous, sound source localisation not only offers insight into localisation, but also extends into the analysis of auditory scene analysis. This may include tasks involving identifying the total number of sound sources

[11] Random noise which, unlike white noise, has a spectral density inversely proportional to the square of its frequency

within a scene, locating multiple or single sources within that scene, and more subjective measures such as envelopment[12] and image width[13]. Three to five sounds of similar loudness and frequency content seems to be the limit at which individual sounds can be identified [58] [59] [60], with human voices generally proving easier to identify in higher numbers than other types of sound. As the number of sources increases beyond this apparent limit, they appear to collapse into what are perceived to be a lesser number of sources within a sound scene, perhaps with a perceived increase in the width of these phantom compound sources.

Spatial separation of sound sources clearly improves localisation accuracy, and complex sounds may be more easily separated by the hearing system allowing for accurate localisation to occur in the presence of larger numbers of sound sources. In terms of localising noise type test signals, participants' have demonstrated difficulty in localising more than three similar sounds simultaneously, perhaps owing to rapid frequency and amplitude fluctuations confounding the ITD cue. Additionally, noise type sounds across similar frequency bands are essentially homogenous and may present the same problem as highly correlating sounds, thus confounding the localisation of multiple simultaneous sources of noise. Should a sound, or small number of sounds within a scene then they may be easier to both identify, and locate. In this case the other sound in the scene may be thought of as interfering sound.

3.2.2 Identifying and Locating Sounds in the Presence of Interference

Although the need to localise multiple sound sources simultaneously may arise in a variety of situations, it is also necessary for humans to be able to identify and/or locate a sound in the presence of other interfering sounds. The ability to isolate a single relevant speech sound, from a background of other similar sounds has been termed the "cocktail party effect" [61]. Although a great deal of literature approaches the topic of perceiving sound amidst interference as a problem of identifying a sound of interest, or

---

[12] A listener's sense of being surrounded by sound
[13] A listener's sense of the scale of an auditory scene or image

understanding it semantically, binaural listening is an integral mechanism with which listeners solve this problem [62].
In perhaps the earliest investigation directly addressing “the cocktail party problem” Cherry [63] proposed several factors that could be used to design a “filter” for separating voices:

- Directional/spatial separation of voices
- Lip reading, gestures, and other visual cues associated with a particular voice
- Differences in voices such as mean pitch, mean speed, etc.
- Differences in accent
- Transition probabilities (based on subject matter, voice dynamics, etc.)

Durlach and Colburn [62], when investigating sound masking levels, found that signal detection against background noise was greatly improved by binaural listening. Where a signal and noise were presented to a both ears simultaneously, where the noise was phase shifted by 180º to one ear, there was a ≈6dB improvement to signal detection thresholds. When the noise was presented in phase to both ears, but the signal was presented with a 180º phase shift at one ear there was a ≈15dB improvement in signal detection thresholds. Such thresholds of signal and masking sound are referred to as binaural masking level differences (BMLD).

Kidd *et al*, [64] conducted a study of binaural masking level difference using 480ms bursts of varying single frequency tones as signal stimuli, comprised of 60ms presentations of eight different frequency bands (with centre frequencies ranging from 215Hz-6112Hz). Interfering/masking sounds were comprised of eight 60ms bursts of either noise[14], or complex multi tonal sound of similar features to the signal stimuli[15]. Sounds were presented form an array of seven loudspeakers, with 30° separation between each. The study concluded that the benefits of spatial separation of a signal and masker

[14] Masking stimuli of this type are commonly referred to as energetic maskers
[15] Masking stimuli of this type are commonly referred to as informational maskers

was dependent upon the type of masking sound. Interestingly Kidd *et al* noted that the greatest improvements in binaural masking level difference (≈30dB) were found for the masking sounds that were similar to the signal stimulus, rather than for the noise masker. In an experiment concerning speech intelligibility, Brungart and Simpson [65] investigated distance, rather than angular separation of the signal and masker sources. In this case, the signal was a randomly selected phrase of recorded speech, and maskers were either another recorded speech phrase of either the same or different gender as the signal speech, or noise filtered to contain energy in the vocal frequency range. It was again found that binaural cues brought about by angular separation of the signal and masker improved speech intelligibility greatly when both stimuli were of the speech type, in the presence of a noise masker the improvement to intelligibility was lessened. Conversely, distance separation of two speech signals did not yield an improvement to intelligibility, but in the presence of noise masking it provided up to 100% improvement.

As mentioned, much work in this area concerns speech intelligibility rather than sound localisation, binaural and distance cues play a role in selective attention to sounds of interest. When humans assess the sound scene in complex sonic environments, it may be safe to assume that neither could be neglected as significant in auditory perception. The improvements in intelligibility originally detected by Durlach and Colburn [62] when phase shifting noise at one ear are translatable into head movement, turning away from a sound of interest would increase the ITD, and thus the phase difference of a sound between the two ears.

3.3 Localisation: Estimating the Distance to Source

Auditory distance perception, the mechanisms by which it is achieved, and its accuracy, have received less scrutiny than aspects of auditory localisation such as direction or altitude estimation. Nevertheless, distance/range estimation is a critical part of auditory localisation [66], and should be considered especially useful for visually impaired individuals, where it may be the only sense of distance to sound source available.

Several potential distance cues have been proposed for scenarios where the sound source and listener are stationary: direct sound intensity, direct to reverberant sound energy ratio, direct and reverberant sound arrival times, changes in spectral content caused by acoustical phenomenon[16], and binaural differences [67]. It has also been noted that errors in distance estimation follow a trend according to the actual range from the listening position to the sound source; for sources in peripersonal[17] space listeners tend to overestimate distance, whereas in extrapersonal[18] space they tend to underestimate them [66] [68] [69]. Furthermore, accuracy in distance perception seems to run contrary to accuracy in localisation in that sound sources appear to be easier to locate in terms of distance when they are placed laterally to the listener, whereas angular localisation accuracy tends to be greatest for sounds placed close to the median plane [70]. This section will explore the complex of cues which contribute to distance perception, and its "uneven" resolution about the listener.

3.3.1 Direct Sound Intensity and Distance Estimation

It is well known that in the free-field[19], sound intensity and distance to sound source are related per the inverse square law:

$$1/R^2$$

Where R is the distance between the sound source and the listening position. As sound pressure is proportional to the square root of intensity then for each doubling of distance between source and listener, a pressure loss of 6dB will be observed.

Direct sound intensity alone however, has proven to be a poor cue for distance estimation: In a classical study of auditory distance perception Georg von Békésky [71] found that in anechoic environments, listeners consistently underestimated distance to human speech sources at distances of 1-10m, observing that underestimation could be as great as a factor of two compared to actual distance. It was however, still demonstrated that as the actual distance to sound source increased, so too did listeners' estimates of distance. Poor

---

[16] Independent from changes in the sound spectrum as it is emitted from the source
[17] The space immediately surrounding the body, usually within reach of the individual
[18] Space outside the reach of the individual
[19] Acoustical environments free from sound reflecting boundaries such as walls

distance estimates were again recorded in a series of experiments by Cochran *et al* [72] in which twenty participants were presented with both recorded, and live speech stimuli in an outdoor environment at distances ranging from 1-29m. Here, underestimates were observed to increase with distance, up to a maximum of ≈30% for the 29m distant source.

Several studies have attempted to determine if, and how, perceived distance change correlates with sound pressure change. In a study of six listeners, Begault [73] reported that an average of -9dB sound pressure change correlated to a perceived doubling of sound source distance, when offered a forced choice between -3, -6, -9, or -12dB. However, values ranging from 8dB [74] and 15dB [75], to 21dB [76] have also been reported necessary for a perceived distance change of a factor of two. Furthermore, sound intensity/pressure loss over distance does not adhere to the previously stated laws when said sound is less than one wavelength from the listening position [77], this means that for simple tones of 20Hz approximately 17m would be required before the previously stated laws began to apply, although natural sounds are generally complex and comprised of multiple audible frequencies of sound neither free field environments[20], nor point sources, are generally found in nature.

Here then is an explanation for the underestimation of distance in extra personal space based upon sound pressure; although pressure is halved (-6dB) for each doubling of distance to a sound source, listeners do not perceive a halving of loudness until the actual sound pressure has been reduced by greater than -6dB, and so accurate perception of sound distance increase based upon intensity alone is compromised. A confounding factor in the use of intensity as a distance cue is that of familiarity; for the intensity of a sound to be meaningful when estimating its distance, the listener must have some concept of how intense the sound is at its source, or for changes in intensity associated with a moving but unfamiliar sound source to give an indication of changing distance. Intensity, therefore, must be viewed as a relative cue, rather than an absolute one [66].

---

[20] At the very least, an open environment would still have a floor from which sound could be reflected

3.3.2 Direct to Reverberant Sound Energy Ratio and Distance Estimation

In their 1934 paper for the Physical Factors Symposium on Wire Transmission of Symphonic Music and its Reproduction of Auditory Perspective, Steinberg and Snow [78] commented that "Depth localisation was found to vary with changes in loudness, the ratio of direct to reverberant sound, or both…". The relationship between direct and reverberant sound energy has long been of interest to the music and audio recording industry. It has also become a point of interest in psychoacoustics where, although it is perhaps less studied than the cues underlying angular sound localisation, its significance as a distance/range cue has not been overlooked.

The direct to reverberant ratio (DRR) has continued to prevail as a topic of study, and a known cue in auditory distance perception, and has been helped by the introduction of binaural audio technology which allows researchers to design and manipulate acoustical environments and the direct-to-reverberant energy ratios associated with them. In a foundational study of direct to reverberant ratio and distance perception Mershon and King [67] presented eighty participants with short noise burst stimuli of high or low intensity (differing by 10dB) from loudspeakers placed at ≈2.7m and ≈5.5m distance in both echoic and anechoic environments. Under these conditions, and after multiple stimuli presentations, estimates of source distance in anechoic environments were underestimated by as much as a factor of 10, however in the presence of reverberation estimates were more accurate. Furthermore, estimates in the echoic environment were stable when compared to those in anechoic environments, where distance estimates became more accurate with increasing presentations of the stimuli. This study demonstrated that although direct sound alone *can* be used to determine the range to a sound source it requires some familiarity, or learning of the stimulus or environment, whereas the presence of reverb allowed estimates to be more consistent without the need for prior experience.

In further studies made for sound sources ranging from 0.6m to 8m in distance, Mershon *et al* [79] [80] reinforced the relative accuracy and stability of distance estimates made in echoic environments vs. those made in anechoic environments. Additionally, in the earlier of the two studies [79], they suggested that no prior knowledge of a given environments acoustics was required for reverberation to act as an absolute distance cue. These results would indicate both the usefulness of reverberation in distance perception, as well as underlining the veracity of the statement that direct (anechoic) sound intensity functions better as a relative cue for distance, rather than an absolute one. The presence of reverb allowed for better absolute distance estimation.

The topic of direct to reverberant ratio and auditory distance perception has also been a subject for study in virtual reality [81], with research examining distance estimation in both binaurally recorded[21] audio, and digitally manipulated audio. In a key study of binaurally recorded audio, Butler *et al* [82] performed a distance estimation test employing low pass, or high pass, filtered noise recorded in echoic and anechoic environments. Participants tended to estimate echoic sound to be at greater distances than their anechoic counterparts, particularly for low pass filtered stimuli. It was also found that localisation ambiguities were less likely to arise in echoic stimuli, although participants were still likely to mistake low pass sounds as originating from behind them. In further studies of binaurally presented audio, Begault [83] found that distance was underestimated by as much as a factor of three for anechoic sound.

Reverberation in binaurally presented audio recordings and renderings is vital for consistent distance estimation, as well as creating the perceptual illusion of sound source direction. Bronkhorst [84] has proposed the following model for predicting distance estimates using direct to reverberant ratio:

$$d_s = A r_h \left(\hat{E}_r / \hat{E}_d\right)^{1/2}$$

where A is constant, r_h is the reverberation radius of the room, and $\hat{E}_r$ and $\hat{E}_d$ are the modified energies of the reverberant and direct sound, respectively. The modified direct

---

[21] Audio that has been recorded with binaural localisation cues encoded into it, by way of a dummy head designed to mimic human physiology and its acoustical impact

energy is obtained by adding the energy of early reflections arriving within the first 6ms to the energy of the direct sound[22]; the modified reverberant energy is simply the total energy of the remaining reflections. In applying the model, the (relative) energies of the direct sound and the reflections are determined using delta functions instead of the individualised HRIRs. This means that the model uses the (single-channel) room impulse

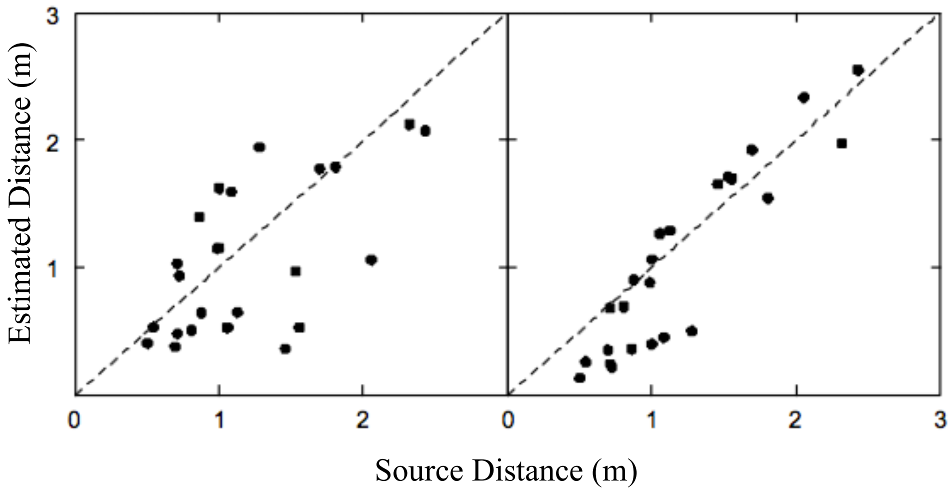


Figure 3.c Graph showing model estimated distance over perceived source distance without accounting for ITDs (left) and in the presence of DRRs with ITD accounted for (right) (Bronkhorst, 2010)

response and not the (two- channel) binaural impulse response as input [84]. Model predictions showed a .62 correlation coefficient with actual distance estimates.

The poor correlation between model predictions and actual distance estimates led Bronkhorst to adjust the model to account for ITD when calculating $\hat{E}_r$ and $\hat{E}_d$ values, such that reflections arriving with ITDs matching that of the direct sound were considered as part of the modified direct sound ($\hat{E}_d$).

The newly modified model resulted in a correlation coefficient of .95 between measured and predicted distance estimates (shown in *figure 3.c*) Indicating that not only is direct to

[22] The model uses a window with a sine-shaped cut-off for this purpose

reverberant ratio significant for distance estimation, but that direct to reverberant ratio may in fact be considered a binaural rather than monaural cue. However, there are thought to be limitations to the maximum auditory distance perceivable in reverberant environments. Described as the auditory horizon, this limit was first theorised in response to the discovery that sound sources located at greater distances in extra-personal space tended to elicit underestimations in perceived distance and that these underestimations eventually became near-constant distance estimations as the sound source was moved even further away from the listener [85] [86]. It has not truly been quantified in terms of a fixed range at which accurate distance perception fails, but it has been shown that the inter-aural cross correlations[23], and temporal and spectral cues thought to be significant in direct-to-reverberant ratio processing become constant at larger distances [87]. Although this would indicate that the collapse of viable reverberant cues is the decisive factor in determining the location of the auditory horizon, it has also been shown that distance underestimation begins before a sound reaches the maximal distances at which distance discrimination should be possible [88] [89].

3.3.3 Spectral Distance Cues

The spectral envelope of propagating sound is a function of distance; for great distances the spectrum varies due to atmospheric absorption of higher frequencies. For close distances, atmospheric absorption may be considered non-significant and instead the spectrum of the sound is modified by the acoustical impact of the head and pinna, creating complex modifications depending upon the distance and angle of incidence of the sound source [90].

Significant modifications to the perceived spectrum of a sound occur because of air humidity, and air movement/turbulence such as wind. These modifications occur over distances greater than ≈15m in 40% humidity air, affecting frequencies greater than ≈10KHz [91]. This dampening of high frequencies occurs in addition to the general sound pressure loss over distance. Several studies have found that low pass filtered sounds tend

---

[23] The measure of similarity of signals received by the two ears

to be estimated as originating from greater distances than unfiltered, or high pass filtered sound, by human listeners [82] [90] [92].

Butler *et al* [82] recorded noise at the ear canals of humans in both anechoic and echoic environments, using low pass (cut-off at 2KHz, 1KHz, or .5kHz), high pass (cut off at 6KHz, 4KHz, or 2kHz) and broadband noise. These recordings were presented to test participants via headphones, where estimates consistently placed the low pass noise at the greatest distance, high pass noise at the least, and broadband noise in the middle of the estimated distance ranges. Little *et al* [92] noted that such changes in spectral shape would not occur naturally as a result of sound source distance, and conducted similar experiments using only low pass filtered broadband noise (with cut-offs of 5KHz, 6KHz, or 6.7 kHz) although reduction in high frequency spectral content was again associated with greater sound source distances this trend in distance estimates was only observed in participants after several trials were completed, suggesting that high frequency loss over distance, much like loudness loss over distance, is a relative cue.

Whilst simple loss of high frequency energy provides some distance cue for distant sound sources, the spectra of nearby sounds and the perception of their distance are, as mentioned earlier in this section, affected by diffraction caused by human physiology [88]. Brungart [93] [94] tested distance judgements of broadband (200Hz-15kHz), high pass (3-15kHz), and low pass (200Hz-3kHz) noise bursts in an anechoic chamber, finding that accurate distance estimates for nearby sounds required spectral content below 3kHz. Kopco and Shin-Cunningham [70] conducted a study focusing on distance estimation for sources located between 15cm and 1.7m of the test participant; noise bursts with centre frequencies ranging between 300Hz-5.7kHz and bandwidths of 200Hz-5.4kHz were presented in a reverberant environment with varying intensity levels to ensure that participants were unable to reliably use the relative cue of sound pressure level when making distance estimates. Distance estimation worsened for both frontal, and lateral sound presentations, as low frequency energy was removed from the stimulus, with the highest performance found for stimuli containing a 300Hz component and it was noted that accuracy was greatest for laterally presented sounds.

Spectral alterations to a sound offer only a partial distance cue as they are only effective cues for sources located at either less than 1m or greater than 15m from the listener [88] [95] as atmospheric attenuation would be perceptually insignificant at less than 15m and, likewise, spectral modifications caused by head diffraction would be insignificant at greater than 1m.

3.3.4 Binaural Distance Cues

Head angle has long been demonstrated to have some influence on distance perception, where listeners are more readily able to judge the distance of sound sources located to the left or right of the median plane compared to sources located at 0° azimuth [70] [96]. Although ITD is independent of source difference, ILD varies as a function of distance for nearby sound sources (located within 1m of the listener) due to the *1/r* law of sound pressure loss over distance. [97]. In both models of nearfield low frequency ILDs where the human head was considered as a sphere [98] and measurements ILDs using recordings made via a KEMAR dummy head [93] [99]) confirmed that ILDs varied as a function of distance. There are however, cases in which binaural psychoacoustic cues have failed to elicit auditory distance perception in subjective trials.

In early studies with single frequency stimulus numerous experiments failed to show any singificant distance perception abilty with binaural cues [100] [101] [102]. Conversely, numerous studies of broadband stimuli have shown that for nearby sources, binaural cues do assist in distance estimation. Holt and Thurlow [96] used thermal noise[24] presented frontally and laterally to test distance estimation and found that estimate accuracy was improved for laterally located sound sources, even when the distance for frontally presented sounds beyond 1.8m could not be determined. When testing with speech stimuli, Cochran *et al* [72] found that head orientation did not affect distance judgement for stimuli presented at greater than 1m. In a similar experiment conducted in an anechoic environment Gardner [103] found that introducing binaural cues through head movement provided only slight benefits to distance estimation for speech stimuli.

---

[24] Broadband sound, similar to white noise

In addition to simple ILD cues, a further potential binaural cue exists in the form of acoustical parallaxing; if monaural cues for the direction of the sound source are available, then discrepancies between the perceived angle relative to the left and right ears may act as a basis for distance estimation, where a greater parallax angle would be observed for nearby sound sources than far ones, increasing interaural differences [104].

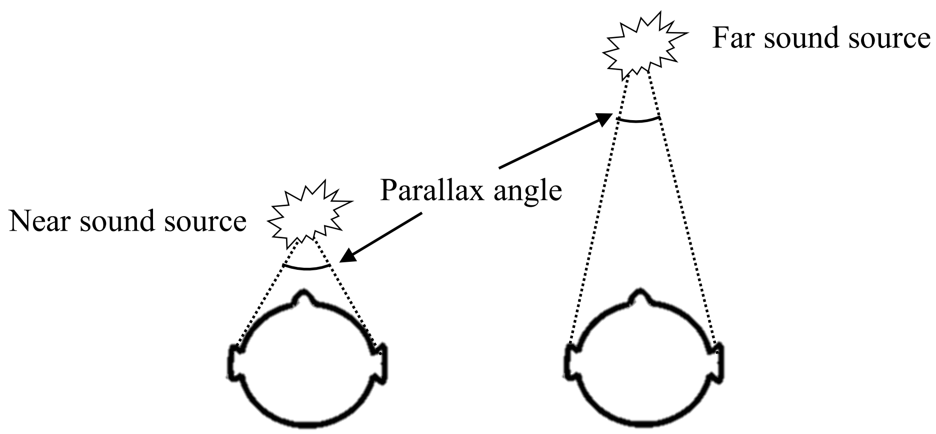


Figure 3.d Diagram showing the larger parallax angle for near sound sources (left) relative to the smaller parallax angle for far sound sources (right) located upon the median plane

Although reliant upon monaural angle estimations at each ear, parallaxing would only be a viable cue when such information was available from both ears simultaneously. Kim *et al* [105] tested this model using synthesized pink noise, with parallax distance cues simulated to eliminate direct to reverberant ratio and other level-based cues. It was found that distance estimates increased with simulated distance until the simulated distance was greater than ≈1m.

Binaural distance cues are certainly viable for near sound sources; however, they appear to be insignificant for simple, stationary stimuli and listening positions, or for sources outside of a 1m radius of the listener when they are the only cue available.

3.3.5 Weighting of Available Distance Cues

Different distance cues become available depending upon factors such as the environment in which a sound source is heard, where the reverberant characteristics will affect the direct to reverberant ratio; the distance and angle of the sound source itself, where different cues are available in near and far sound sources, and where laterally placed sound sources may present binaural cues; and the spectral content of the sound in question, where certain cues such as atmospheric dampening may become unreliable due to a lack of high frequency content in the original sound, the reverberant properties of the environment may vary over frequency, and simple sound sources may confound binaural cues.

Auditory distance perception must, therefore, be achieved by combining these cues depending upon which are available in each situation. Kopco and Shin-Cunningham [70] suggested that direct to reverberant ratio and ILD information may be selectively combined, facilitating accurate distance estimation in reverberant environments. Although offered as a reasonable explanation for the results found during their experiments, the phenomenon remained largely untested in their work. In a 2002 adaptive study of perceptual thresholds in for distance cues in virtual environments, Zahorik [106] stated that pressure level and direct to reverberant ratio cues were weighted flexibly, where (if available) the relative intensity cue would lead to more precise distance estimation and conversely the coarser cue of direct to reverberant ratio would be used in situations where the relative cue was unavailable. Furthermore, Zahorik asserted that the perceptual basis for direct to reverberant ratio calculation was unlikely to be time-based, as the onset and duration of stimuli did not have any apparent effect on distance estimation using direct to reverberant ratio alone. Instead it was suggested that spectral envelope changes, or variations in direct to reverberant ratio bringing about changes in the spatial qualities of a sound, would be a more salient basis for direct to reverberant ratio calculation. This would indicate that whilst direct to reverberant ratio is a vital cue for absolute distance estimation, relative cues such as intensity level, and ILD would take precedence for distance perception whenever available.

In research combining an exploration of sound movement perception in anechoic and reverberant space, with ecologically valid sounds[25], with or without interfering sound sources, Lundbeck *et al* [107] employed a methodology which explored both angular movement about the listener as well as linear changes of distance between sound source and listener. It was found that, in conditions with greater numbers of interfering sources, the addition of reverberation created a slight but statistically insignificant improvement in minimum audible movement distance thresholds, whilst minimum audible movement angle thresholds were always somewhat worsened in the presence of reverberation. The slight improvement in distance movement perception thresholds was attributed to a complex of cues including monaural spectral cues, direct to reverberant energy ratio. A note of caution was offered by the authors, in that the use of non-individualised HRTFs could have led to reverberation causing improvements in measured thresholds as it could ameliorate certain problems with binaural audio presentation (these issues are discussed further in *Chapter 4.2*).

The work of Lundbeck *et al* [107] drew interesting comparisons between individuals with healthy hearing, and those with age-related hearing loss. Whilst it is thoroughly sensible to predict differences in comparisons of the auditory acuity of these two groups, the auditory perception of visually impaired and sighted individuals has also been a topic of study for some time, where comparisons with sighted, early-onset, and late-onset sight loss have revealed certain differences between groups.

3.4 Perceptual Studies of the Early & Late Blind Visually Impaired

The relationship between spatial hearing acuity and vision loss is complex. In the absence of eyesight, auditory perception is the only major sense capable of providing information about the environment beyond peri-personal. As detailed previously, the cues used by the auditory system to determine spatial information vary in saliency and usefulness. Vision loss can lead to distortions in an individual's mental representations of space (as detailed in *4.4*) but has also been linked to alterations in spatial auditory perception. This already

[25] A telephone, a fountain, a bell, and a drink being poured

complex issue has been complicated yet further by the fact that differences in auditory perception may arise depending upon the age at which an individual is deprived of vision.

3.4.1 Some Distinctions Between Late and Early Blindness

The possibility that differences in some aspects of auditory perception may arise in the visually impaired depending upon their age at the onset of sight loss was first explored by Hubel and Wiesel who, following a period of study on the effects of monocular sight deprivation in kittens [108] [109], in 1970 investigated what they referred to as "the period of susceptibility" to neurophysiological changes in response to sight deprivation [110]. They found that temporarily depriving a kitten of eyesight in a single eye for as little as three days was enough to trigger a decline in the proportion of cells in the primary visual cortex that responded to the deprived eye. This susceptibility was noted to decline over the first three months of the kitten's life whilst adult cats showed little susceptibility at all, even to longer periods of sight deprivation. These studies of monocular sight deprivation where the first to show that even temporary sensory deprivation in the immature brain could lead to long lasting changes in brain function.

Subsequent advances in neuroimaging techniques have allowed research into the living brain to thrive. The visually impaired have proven to be effective models for studying the plasticity of the human brain [111] where the distinction between early onset and late-onset blindness has been broadly defined in terms of two time windows:

"**Early blindness**: refers to cases of blindness that occurred during the first few years of life, generally prior to the age of 5. However, there are multiple exceptions, with some studies including participants up to 14 years of age in what are defined as early blind groups. Also, early blind groups often include congenitally blind individuals, unless otherwise specifically stated. While early and congenitally blind individuals are often pooled together, more recent studies have started segregating them into separate groups, as even a few years of visual experience could strongly alter the functioning and the anatomy of visual structures.

Late blindness: generally, refers to cases of blindness that began after puberty (typically >16 years of age) or in adulthood. Again, there are exceptions to this with some studies including individuals with ages of onset as low as 7 years of age.

The lack of consistency in defining blind groups across studies has had two major consequences. The first is the often omittance of individuals with intermediate onsets of blindness (e.g., between 5 and 16 years of age), which of course introduces a strong sampling bias when attempting to relate the age of onset of blindness to a behavioural or neuroanatomical measure. The second is the undesired overlap between defined groups from different studies, where a given individual would be considered as 'early blind' in one and as 'late blind' in others." [112]

The continued neurological assessment of early blindness in cats has revealed fascinating insights into the plasticity of the brain; it has been demonstrated that not only do regions normally associated with vision begin to respond instead to auditory stimuli [113] but that neurons outside of the normal auditory regions may become tuned to auditory spatial information [114].

As may be expected, such topics of study have also expanded into human research. Although standard audiometric thresholds have not been found to differ between sighted and visually impaired humans [115] [116] certain auditory perceptual differences have been measured between early blind, and late blind and sighted individuals.

3.4.2 Auditory Spatial Tuning in Early Blind, and Late Blind and Sighted Individuals

Although a relatively large number of perceptual studies of the sighted and visually impaired have been conducted in the past, fewer have been conducted to compare late and early blindness. Early experiments comparing the auditory spatial abilities of blind and sighted individuals tended to focus on the congenitally blind and found that, when blindfolded, sighted individuals performed similarly to the visually impaired. For example, in a 1964 study of sighted and blind individuals [117], Fisher tested the spatial awareness of participants to both auditory and tactile stimuli. The auditory component of the test consisted of white noise played at different locations across a 180° horizontal arc in the frontal plane. When locating the test position with either manual pointing, or turning to face its perceived location, no significant differences were found in the performance of either blind or sighted participants. Zwiers *et al* [118] again found that azimuthal sound localisation was identical for blind and sighted participants, and that in the presence of interfering noise the performance of both groups was similarly affected by worsening signal-to-noise ratios. As with the work of Fisher, this study was limited to stimulus presented in the frontal plane. In this case however, the spatial extent of sound presentation was limited to within 50°. Studies limited to frontal sound presentations at fixed distances are perhaps less likely to reveal auditory perceptual differences between blind and sighted individuals, as many later studies in which a difference *has* been measured have not limited their test stimuli in such ways [119] [120] [121]. This may well be due to the apparent bias in “tuning” of attention in the auditory system to frontally located stimuli in sighted humans [122]- it is possible that auditory spatial discrimination approaches some maximum threshold in both sighted and blind individuals when dealing with frontally located sounds.

The late blind offered an interesting opportunity to further study differences between the sighted and visually impaired; if some differences exist between the early/congenitally blind and sighted, what about those visually impaired individuals with early-life visual experience? In a neurological study of late and congenitally blind, and sighted individuals Collignon *et al* [123] paired a functional magnetic resonance imaging with auditory task performance. The tasks consisted of a spatial differentiation and pitch differentiation threshold task. For pitch variation, a 1kHz reference tone was played first, followed by the test tone. The frequency of the second sound was incrementally changed by steps of five cents[26] compared to the first. For spatial variation, the second sound was adjusted in terms of azimuth location by simultaneously increasing the ILD by steps of 0.2% and the ITD by steps of 20µs compared to the first sound (located at 0° azimuth). In both pitch and localisation tasks, both the pitch and azimuth location of the sound were changed, however participants were instructed to report either only the location of the second sound (to the left or right of the first) or the pitch (either higher or lower than the first). Given the frontally clustered presentation of the sounds, it is perhaps no surprise that all groups showed highly similar behavioural performance, more interesting were the result of the neuro-imaging carried out during the tasks.

[26] A logarithmic unit of measurement of musical intervals

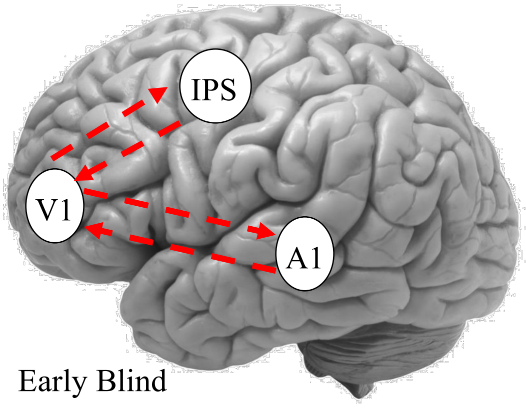


Figure 3.e Schematic representation of auditory information flow from the primary auditory cortex (A1) toward the primary visual cortex (V1) including the intraparietal sulcus (IPS) in early and late onset blind participants (after Collignon et al [76])

These tests revealed a difference in neurological organisation between early and late blind individuals. Although both groups recruited occipital regions for auditory processing, only the early blind group showed activation of the middle occipital gyrus and cuneus for spatial processing, supporting the notion that a critical period of spatial tuning for occipital regions must exist in humans. Collignon *et al* [123] concluded that a difference in connection between auditory and (normally) visual cortices of the brain in early blind individuals was the best explanation of their results (a schematic based upon this model is shown in *figure 3.e*).

3.4.3 Spatial Discrimination in Early and Late Blind Visually Impaired Individuals

Considering the neurological differences between early and late blind individuals, it would seem a safe assumption that some functional differences in auditory spatial perception should arise. Although many studies have found similarities between the two groups [117] [118], others have found differences of both improved and worsened performance in the early blind when compared to late blind, and even sighted groups.

In a study of auditory motion, Finocchietti *et al* [124] found that sighted, and late and early blind individuals were correctly able to determine and recreate the trajectory of a moving sound source within a 45cm radius centred at 0° azimuth and at arm's length of each participant with similar levels of accuracy, except where those movements occurred below the horizontal plane of the head [124]. Here the early blind group showed error margins approximately four times greater than those of their sighted and late blind counterparts. Although completing this task involves not only auditory spatial perception but (owing to its design) aspects of proprioception and memory as well, it does to some extent echo the findings of more audio focused experiments. In a binaural localisation study, including interfering noise, Zwiers *et al* [118] found that elevation discrimination in early blind participants deteriorated rapidly in the presence of increasing noise, when compared to the sighted group. This led to the postulation that monaural spectral cues were masked by the noise, confounding the blind group's vertical localisation ability. Voss *et al* then studied the monaural localisation abilities of early blind and sighted participants by occluding one ear of each individual during tests [125]. They found that although horizontal plane localisation was better for the early blind group under this condition, vertical plane localisation was still worse, an indication that perhaps loss of monaural information was not the single cause of worsened performance in Zwiers *et al's* earlier study. Interestingly this worsened performance was pronounced in the vertical plane *above* the horizontal plane rather than below (as in the results of Finocchietti *et al.*) and there was an inverse relationship between horizontal and vertical performance: for blind individuals, greater performance in horizontal localisation correlated with worsened performance in the vertical plane. Although these reports of worsened performance in early blind groups conflict in some areas, they do share a common theme; when sound is in the frontal hemi field, horizontal localisation is similar in all people tested, however the early blind do appear to underperform when identifying elevation or tracking vertically displaced sounds.

When test stimuli are moved beyond the confines of fixed ranges in the frontal region, some increased auditory abilities may be observed in early, and sometimes late, blind individuals. For instance, Lerens and Renier [126] found that early-blind individuals were able to discriminate target sounds from distracting interference faster than their sighted and late-blind counterparts, noting that auditory attention appeared to be biased towards frontally located sounds in sighted groups- a bias which did not appear to exist in the early-blind [127]. In a study of minimum audible angle, and minimum audible distance changes in both frontal and peripheral space, Voss *et al* [128] found that whilst there were no minimum audible angle differences between early-, late-blind and sighted participants for frontally located sounds, the early-blind group showed improved performance to the other groups for peripheral sound sources. Both early-, and late-blind people outperformed the sighted group for sounds located to the rear of the listening position. They also found that both blind groups outperformed the sighted group for minimum audible distance changes in sounds located at 0° azimuth. These findings support the earlier assertions of Röder *et al* who measured improved auditory spatial perception for peripheral sound sources in the early-blind, accompanied by differences in the scalp distribution of brain activity between early-blind and sighted participants [120] which itself was perhaps an early indication of the existence of the neurological differences which would later be modelled by Collignon *et al* [123] (shown in *figure 3.e*). It has since been found that for stimuli approaching the frontal plane, sighted individuals display the most accurate sensitivity to stimuli location changes when using a combination of auditory and peripheral visual senses rather than either sense alone [129].

3.4.3 The Implications of Inequality in Auditory Perception

Despite there being no implied or measured difference between what a sighted or visually impaired humans hear; a strong case has been made for differences in the neural response to sound. Although the visually impaired experience no improvement to auditory perception for frontally located sounds, indeed they may in fact underperform in elevation based sound localisation tests when compared to sighted individuals, there is evidence to suggest that sound localisation is improved for rear and peripherally located sounds.

The notion that late-blind individuals perform similarly to their sighted counterparts in auditory perception tests is supported by measured differences in brain activity during localisation tasks. Many perceptual studies compare early-blind with sighted participants however, there are less studies which compare late- and early-onset blind participants. Pooling data from these groups is sometimes troublesome, as noted by Voss [112] there is a lack of consistency in the way in which these groups are formed with a lack of clarity as to the upper age at which at visually impaired individual might be considered early-blind.

It would be prudent to distinguish between these two blind groups in future studies of auditory perception, and the present study of behaviour and auditory perception should be no exception- it might be the case that the visually impaired move their heads in a different manner, or place different levels of importance upon their head movement, compared to their sighted counterparts, certainly there is some evidence to believe that differences may in fact occur due to reduced head movement for the purposes of posture control [14]. For instance, sound locations which may be ambiguous to a sighted person and therefore elicit some head movement in an effort to either visually or audibly determine their whereabouts, may in fact not seem ambiguous to a visually impaired individual owing to their apparently improved auditory spatial acuity in some regions. But, if the late-blind have comparable auditory perception to the sighted then behaviour between these groups may not differ.

4. The State of the Art in Binaural Virtual Reality

This chapter may be considered both a background to the aims of this thesis, and an extension of the literature review. It outlines the implementation of binaural auditory virtual reality and its potential applications, also highlighting the debate regarding the implementation of head tracking in such systems, and whether it enhances auditory localisation and/or navigation in virtual environments, which may in turn facilitate complex tasks such as the formation of mental maps.

Although, as described in *chapter 3*, many early experiments and studies in virtual reality were conducted with the aim of investigating some aspect of human auditory perception or creating training simulations for hazardous activities [130], systems are now also commercially available for entertainment purposes.

Running concurrent to the afore mentioned developments in virtual reality, there has long been an interest in its application as an assistive technology for the visually impaired; from early experiments to simply translate computer visual interfaces into auditory ones [131] to efforts to create explorable auditory environments [8] and portable systems for navigation and wayfinding [132]. To create an immersive audio world in virtual reality that is congruent with real world environments it is necessary to devise a system that allows the listeners to perceive sound in three dimensions. It must be possible for a virtual reality user to hear where a sound is coming from in the virtual space around them, that this sound must be perceived to originate in any direction and distance relative to the listening position- a feat not possible with traditional stereophonic or surround sound[27] audio reproduction.

Although systems using multiple speakers placed around the listener (such as Ambisonic speaker arrays, in which four, nine, or more speakers [133] are used to create a multi-dimensional sound-field about the listening position) are available, current trends in

---

[27] Such as 5.1, or 7.1 speaker systems

virtual reality research point towards a preference for binaural rendering over stereo headphones [134] when creating virtual auditory displays (VADs).

There are numerous ongoing debates surrounding the application of binaural audio rendering in virtual reality. Although many of these debates focus on use cases for the general population it worth considering applications for the visually impaired as a separate and special case due to the end user's reliance upon auditory perception for experiencing. These debates, and a contrasting summary of general directions in auditory virtual reality and for the blind are presented in this chapter, along with a broader background on the implementation of auditory virtual reality.

4.1 Binaural Audio Rendering

Binaural rendering allows for the listener to experience a detailed, spatialised, auditory scene by imposing the psychoacoustic cues (discussed in *chapter 3*) over monaural[28] audio signals. This section discusses the appropriate measurement of localisation cues, and their implementation in binaural audio rendering.

4.1.1 Applying Binaural Cues for Spatialising Audio: Convolution

Convolution is a mathematical operation on two "input" functions which produces a third "output" function. In digital audio terms, both input functions are digital audio signals, and the output function is a third digital audio signal combining certain properties of the inputs.

Convolution is particularly useful when the properties of a complex system must be simulated, as it allows for this simulation to take place without the need for a model of the system to be developed. Convolution is extensively used in audio signal processing where the desired acoustical properties of an environment must be imposed upon a signal recorded in a space lacking those properties. In this case, the desired acoustical environment represents the system, and the signal to be manipulated represents one of the input functions, the second input function takes the form of another audio signal which

[28] Single channel audio, with no original stereo, or spatial properties

is a measurement of the system environment, known as an impulse response (IR). After convolution is applied to these signals the output function will represent the original audio signal with the desired acoustical properties applied to it, a process known as convolution reverb.

In the digital domain, convolution of two signals is simplified by the fact that these signals are represented numerically such that convolution can be applied by multiplying samples in each of the inputs and adding the products together per:

$$y(n) = \sum_{m=0}^{N-1} h(n-m)x(m), \qquad n = 0,1,2,\ldots$$

Which is constructed programmatically thus:

$$y_n = (x_n \cdot h_1) + (x_{n-1} \cdot h_2) + (x_{n-2} \cdot h_3) + \cdots + (x_{n-m} \cdot h_m)$$

And commonly denoted by the following:

$$y = x \otimes h$$

Where *y* is the output signal, *x* is the signal to be manipulated, *h* is the measurement signal, and *m* is the final sampled value in the input signals. This offers fast convolution [135] which can be achieved in real-time, allowing for an output signal to be computed whilst being listened to.

4.1.3 Applying Binaural Cues for Spatialising Audio: Impulse Responses

Successfully achieving digital signal convolution requires the use of a suitable impulse response, which can only be measured for a linear and time invariant system; the system's output response to a sum of inputs is equal to the sum of outputs produced for each input applied individually, and that system's properties and parameters are constant over time.

This means that a suitable measurement of the system can be obtained by applying an impulsive input to it and recording its output response.

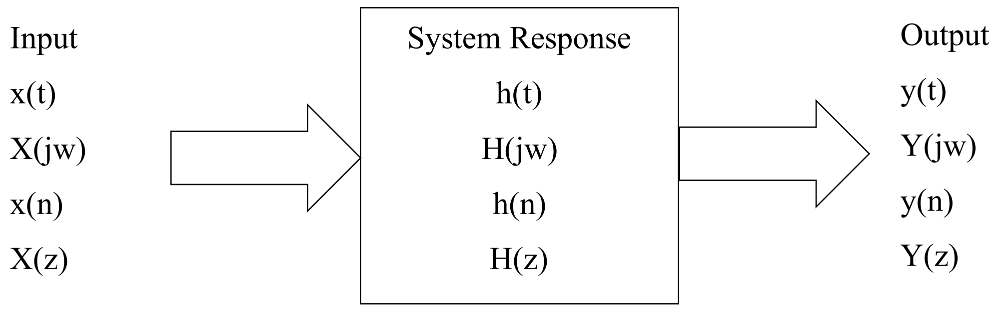


Figure 4.a A linear, time invariant system described using linear transform, after Riederer [223]

The ideal input for such measurements is the Dirac δ, a theoretical function that can be thought of as an infinitely high, infinitely thin peak with an area of one [136]:

$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

The advantage of the Dirac δ signal is that, being infinitely short, it should be relatively simple to distinguish between the original impulse and the system's impulse response at

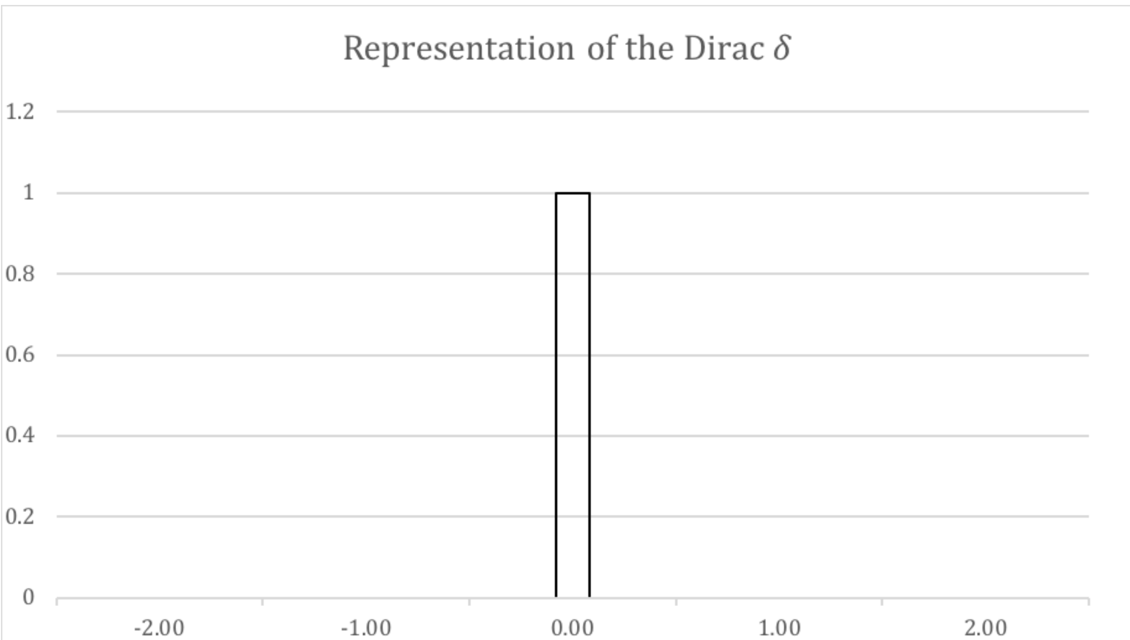


Figure 4.b A representation of the Dirac δ which can be thought of as a function which is 0 everywhere on the real line, except for at the origin

the output. In practice, however, achieving an acoustical measurement (such as a head-related impulse response) using an infinitely short signal is impossible.

Approaching the criteria of the Dirac δ in acoustical measurements can be achieved in several ways, of which four common methods are discussed below:

- Approximated Impulse Response (AIR): this method attempts to utilise a close facsimile of the Dirac δ by using a short, burst of broadband noise as an impulse [137]. This noise is often created mechanically, with a blank firing pistol for example [138]. This can lead to problems with repeatability, particularly in terms of the directional sound power of the source.
- Maximum Length Sequence (MLS): this method uses a computer generated, periodic, pseudorandom noise with an ideally flat spectrum over one period [139]. The measured impulse response of the system is obtained through circular cross-correlation between the original input signal from the recorded output of the system. This yields a periodic impulse response $y'[n]$ which relates to the linear impulse response per:

$$y'[n] = \sum_{l=-\infty}^{+\infty} y[n + lL]$$

Where L is the length of one period. An error, known as time-aliasing, arises with this technique when the length of one period is shorter than the length of the impulse response to be measured. Distortion can also arise in the measurement, because of non-linearity in the measurement equipment, particularly the loudspeaker from which the maximum length sequence may be played [140].

- Time Stretched Pulse (TSP): this method uses a time expanded, computer generated impulse signal with a flat spectrum [141] which is intended to yield a high signal to noise ratio, and nominal distortion from measurement equipment non-linearity by increasing the amount of sound power emitted for a fixed magnitude of signal. To obtain the impulse response, a time compression filter must be used on the measured output to reduce it to a non-time stretched signal.

Although distortion is reduced when compared with that of the maximum length sequence, it is still present within the time stretched pulse impulse response [142].

- Logarithmic Sine Sweep Technique (LSST): this method, uses an exponential time-growing frequency sweep to obtain a linear impulse response, and isolate the impulse response from distortions caused by the measurement equipment. This is possible because the linear impulse response, and distortions are de-convolved simultaneously [143].

The logarithmic sweep is described by:

$$x(t) = \sin\left[\frac{T\omega_1}{\ln\left(\frac{\omega_1}{\omega_2}\right)}\left(e^{\frac{t}{T}\ln\left(\frac{\omega_1}{\omega_2}\right)} - 1\right)\right]$$

Where ω_1 is the initial radian frequency and ω_2 is the final radian frequency of the sweep of duration *T*. The deconvolution process is achieved through the linear convolution of the system output *y(t)* with an inverse filter *f(t)*:

$$h(t) = y(t) \otimes f(t)$$

The inverse filter *f(t)* is described by:

$$x(t) \otimes f(t) = \delta(t - K)$$

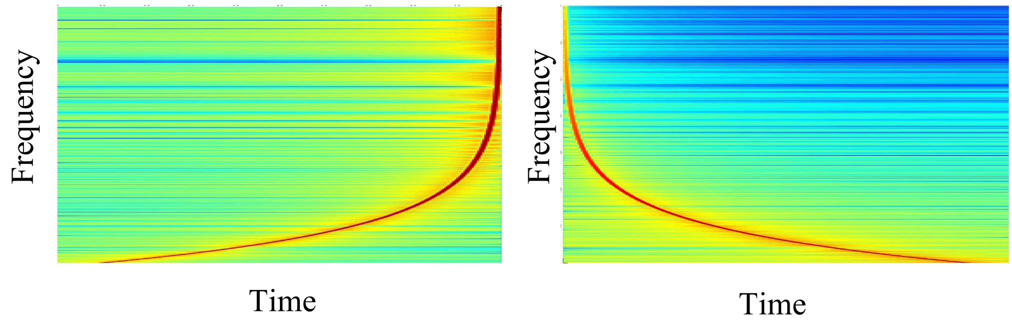


Figure 4.c Spectrograms illustrating the logarithmic sine sweep (left) and inverse filter (right)

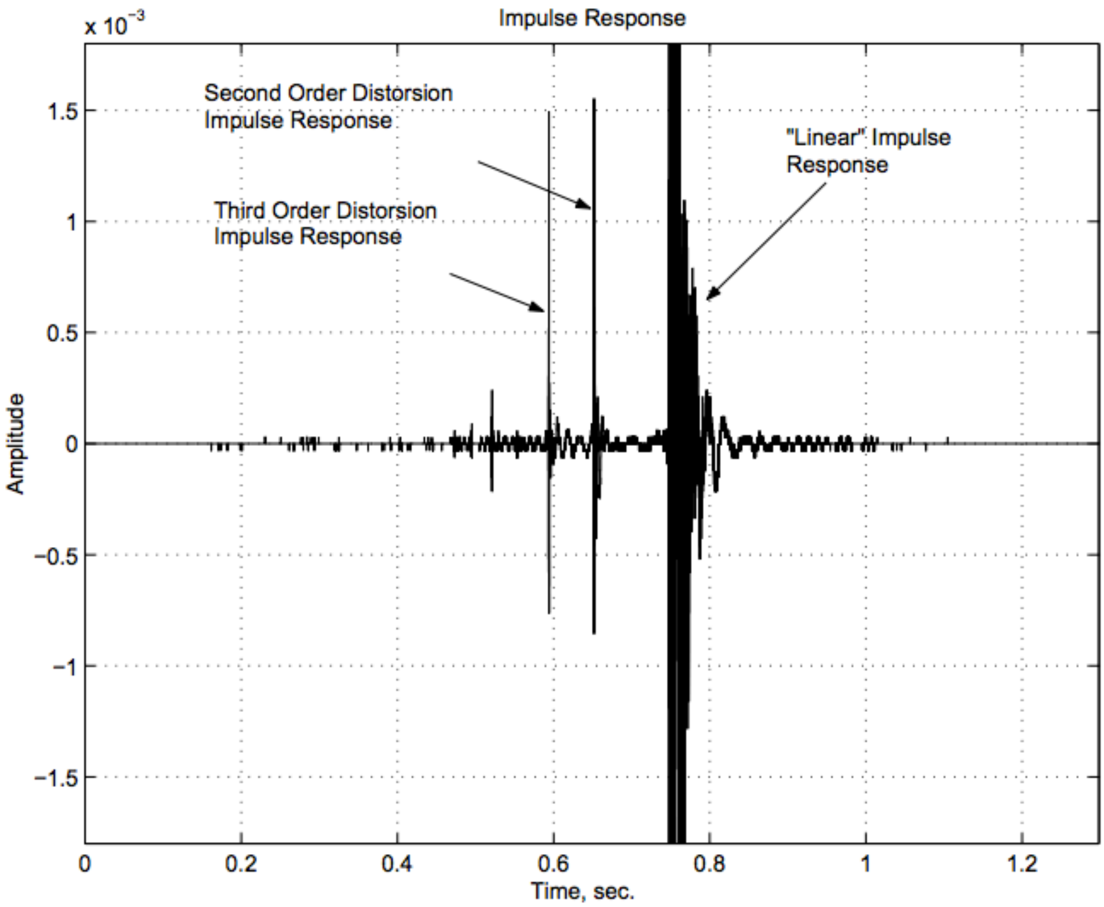


Figure 4.d Impulse Response measurement using the LSST with linear impulse response and distortion annotated, after Stan et al (2002)

The resulting output signal after deconvolution is a linear impulse response, with distortion artefacts appearing before the actual impulse response.

Although the logarithmic sine sweep technique offers a distinct advantage in terms of its ability to reject distortion in measurements, it is worth noting that approximated impulse responses requires no additional post-processing to obtain an impulse response and is therefore the simplest of these techniques, whilst the input signal of the maximum length sequence and time stretched pulse may be shorter than that of the logarithmic sine sweep technique for audible range measurement. Therefore, there are practical circumstances under which techniques other than the logarithmic sine sweep technique may be preferable. The long duration of the sine sweep does offer one great advantage over the approximated impulse response technique, the sweep is typically several seconds long and should therefore be appropriate to capture pinna related spectral notches which arise after the first .5ms of exposure to incoming sound, as measured by [55]. This would seem to be an ideal technique for capturing head/body related measurements for binaural rendering, as it offers advantages in terms of both technical requirements and the validity of psychoacoustical cues advantages.

4.1.4 Applying Binaural Cues for Spatialising Audio: Cue Measurements

Digitally recorded monaural audio can be manipulated through the process of convolution to include the psychoacoustic cues necessary for it to be perceived as having originated from a point in space about the listener. To achieve this a set of measurements; head related impulse responses (HRIRs) and the frequency domain equivalent head related transfer function (HRTF) is required.

Although there are commercially available systems for head-related impulse response measurement [144] most measurements and measurement protocols are made as part of ongoing research into virtual reality, psychoacoustics, and spatial audio reproduction [145] [146] [147]. The general principles of head-related transfer function measurement are that a subject (either a dummy head, or human) is measured acoustically such that the physiological impact upon a sound wave arriving from any given angle about the head is captured. To achieve, this multiple measurements of the subject are made and compiled into a library of head related impulse responses (HRIRs), each head-related impulse response representing the sound recorded at the ear canals, modified by the subject's physiology as it arrives at the head form a specific angle.

Although measurement methodologies vary, particularly in terms of the angular resolution (number of discrete angels measured), the sound type/source and microphones, and whether (in the case of methods using a dummy) the measurements include a torso as well as head, a general outline of measurement protocols based upon the IRCAM [146] method serves to illustrate the process well:

1. The subject is placed in an anechoic environment (thus rejecting room acoustical properties form the measurements)
2. a microphone placed at the entrance to each ear canal such that the ear canal is completely occluded (thus rejecting ear canal resonance from the measurements)
3. A loudspeaker is placed on a mobile platform (such as a small crane), such that it's angle and distance from the subject is adjustable
4. A test signal is played by the loudspeaker and simultaneously recorded by the microphones at the subject's ear canals
5. The position of the speaker is adjusted in terms of relative angle to the head, or distance from the head
6. Steps 4 and 5 are repeated until the desired angles, and distances, of measurement have been obtained
7. The resulting impulse responses, in the form of audio recordings, represent the desired HRIRs of the subject, and are free from room acoustical information, and ear canal resonance information

A vital factor in the measurement of HRIRs and creation of databases is the nature of the measurement signal and the resultant measurement that is recorded: since the goal of such measurements is not only analysis, but application in audio rendering via convolution, the measurements must be recorded in a form suitable for convolution processing (as discussed in *chapter 4.1.2).*

The process of convolution is extremely useful in creating spatialized audio, as it allows for measurements of a complex system, such as the acoustical properties of the human head, to be applied without the need for modelling of the system. Similar procedures involving a microphone and loudspeaker placed within an empty room can also be used to create impulse responses of that room's acoustical reverberance. Once such measurements are made, monaural audio signals can be convolved with the necessary impulse responses to create convincing virtual sound scenes, which can be experience in 3D by the user.

4.1.5 A Schematic Model for Combining Impulse Responses and Audio in VR

To illustrate the application of the techniques discussed previously in *chapter 4.1.1-4* a hypothetical model for combining impulse response measurements of the human head and acoustical environments with monaural audio signals is presented here. This model assumes that spatialized 3D audio, environmental acoustics, and user movement/control are required for auditory - although the necessity and benefit of the latter 2 features may be debatable (and is discussed from *chapter 4.2* onwards) it is useful to demonstrate a means by which such a virtual reality system might be achieved, and the complexities that the creation of such a system entails.

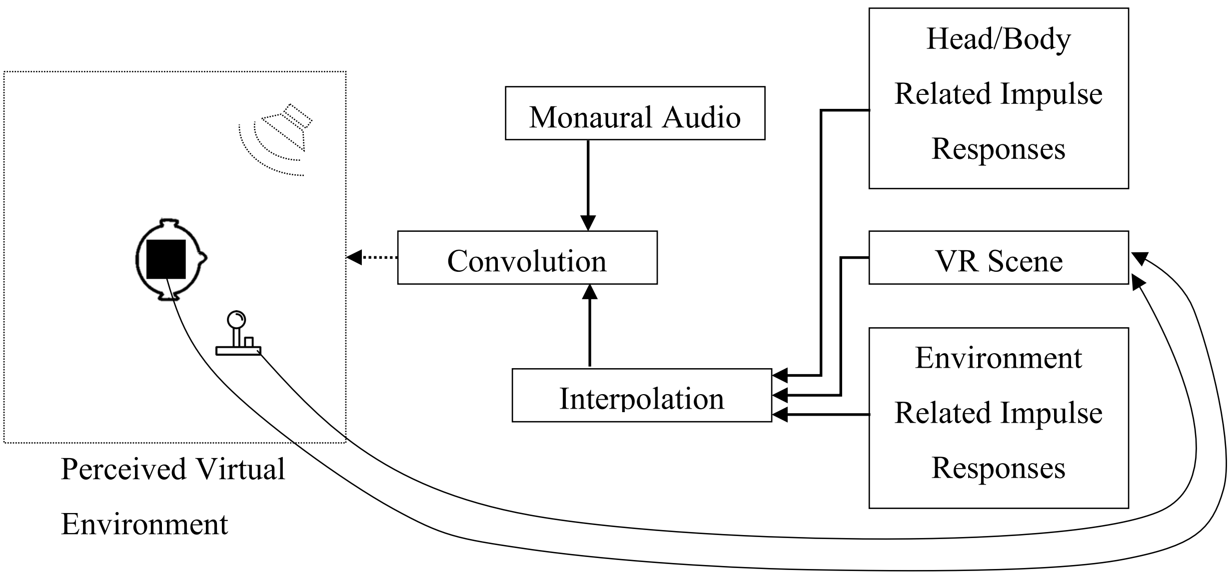


Figure 4.e Schematic representation of an auditory virtual reality system with binaural rendering, room acoustical emulation, and joystick and head tracker user controls

The system in *figure 4.e* shows that to create a meaningful virtual environment, a calculated virtual scene containing all the desired sound emitting and reflecting objects must be combined with the appropriate impulse responses to cause the user to perceive them as correctly placed within the environment. The virtual scene may be considered as a database of information regarding the target environment to be created, including the position of sound sources and their associated monaural audio, as well as the acoustically relevant properties of the room such as the position of sound reflecting surfaces.
The inclusion of head tracking in the system described by *figure 4.e* adds a level of complexity yet undiscussed in this text- that of the body related impulse response (BRIR). The methods discussed in *chapter 4.1.4* assume that the head and body of the subject, and therefore the emulated head and body of the end user in virtual reality, remain static relative to one another. In systems where movements such as head turning are emulated, this is not the case- when the user turns their head alone the position of the shoulders and upper torso, and any sound reflections they elicit, will be in a different location relative to the head and ears. To accurately measure such changes, it is necessary to modify the measurement methods described in *4.1.4* by adding an additional phase of measurements in which the head, rather than the loudspeaker, is moved about its fully range of motion relative to the position of the subject's body.

The required impulse responses are convolved with the monaural audio, after any HRIRs not included in the original measurement set have been interpolated (discussed further in chapter *4.2.1*). Should the user's position, orientation, or head angle change relative to the environment or sound sources within it then, by way of selecting new impulse response measurements for convolution, the scene must be updated such that it appears to remain consistent with the user's new posture (discussed further in chapter *4.2.-3)*. The complexity of the scene in terms of both number of sound sources and the faithfulness of the environmental acoustics, as well as the level of control that the user has over their emulated movements within it, contribute directly to the computational resources required by the system [148].

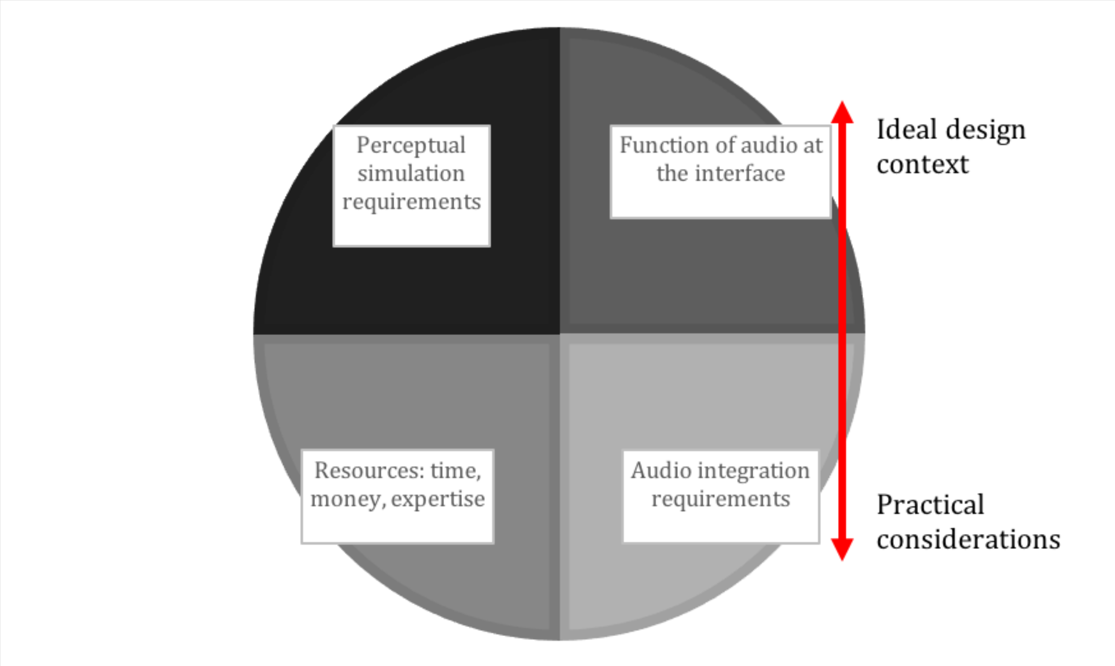


Figure 4.f A requirements versus resource allocation chart for designing audio subsystems within virtual reality or multimedia, illustrating the potential conflict between achieving an ideal design whilst working within technical and financial limitations [14]

The potential complexity and expense of systems prompted Begault [9] to echo the earlier calls of Wenzel [149] to consider what is required by a system's user in order to be fit for purpose, offering the model in *figure 4.f* to represent the factors involved in balancing the needs of a system versus the resources available with which to achieve it.

It would therefore be prudent to exam the current issues in creating binaural auditory virtual reality (see *4.2*), with a view to considering which features are likely to be the most significant in an assistive system for the blind.

4.2 Current Issues in Binaural Rendering for Virtual Reality

This section discusses some current issues in binaurally rendered 3D audio; the implementation of, and concerns surrounding, the use of head-related transfer function/head-related impulse response for spatializing sound; the necessity and accuracy of emulated user head movements, and the practical considerations of implementing of emulated user movements in a virtual environment.

A few critical problems are commonly encountered when implementing binaural audio [5]:

- Front-to-back reversals; a phenomenon where sounds are perceived as having originated from the opposite side of the frontal plane to that intended [150]
- In-head-localisation/listening; often equated to severe misperception of distance, users perceive sounds as originating from within their own head, rather than in the environment around them [4]
- Poor localisation performance; where sounds are not perceived as originating from their intended location, but the angular misperception of the sound source is not great enough to constitute front-to-back reversal [151]

Although strictly speaking these are issues of implementation; they are caused by what may generally be compromises in the calibration or implementation of some part of a binaural rendering system, though they are often measured by evaluating user performance or subjective experiences within said system [151] [17]. As may be seen in the following sub-sections, the guiding principle of much of the work reviewed here seems to be that an effective binaural VR system would elicit similar performance from its users in virtual auditory tests as have been observed in either non-virtual control group tasks or previous studies conducted in the real world (such as those presented in *chapter 3*).

4.2.1 HRIR/HRTF Measurement and User Compatibility

As described in *4.1.4* the measurement of HRIRs and the associated HRTFs is achieved by taking measurements in discrete steps around the surface of a sphere about the subject's head (either that of a dummy, or human participant). To use such measurements in a way that ensures adequate sound localisation performance, it is necessary to ensure that a smooth transition between each discrete measurement is possible, such that the user perceives a cohesive transition in directional cues for sound rendered in the environment [152]. For this reason, measurements must either be made at steps below the minimum audible angle, or interpolation between measurements made in larger steps must be implemented [148].

Since the human hearing system has a theoretical spatial resolution of $\approx 1^o$, or better, for certain frontally located sound sources [153] interpolation is a favourable option compared to making discrete measurements for each required angular position. Langendijk & Bronkhorst [154] found that HRTFs measured with 5.6° resolution in the horizontal plane and completed with interpolation could not be distinguished from fully measured head-related transfer function sets. In a paper reporting on a novel head-related transfer function measurement apparatus Hosoe *et al* [155] made measurements in 5^o steps on the horizontal plane, and ≈6° steps in elevation, yielding a total of 2,088 measurements whilst still not achieving the maximum spatial resolution of the human auditory system (which, without interpolation, would require approximately five times as many measurements on the frontal horizontal plane alone).

Pollow *et al* [156] have categorised the interpolation of HRTFs into 2 main types:

- Local interpolation (LI); using neighbouring measurements for calculations, which offers relatively fast calculations [154] [157]
- Global interpolation (GI); using the entire measurement set for calculations, which offers closer agreement with measured data [158] [159] [160]

Although a practical case can be made for using partially interpolated data to achieve binaural rendering, the relative success of spatial audio has also been shown to depend upon how well the measured data fits any given user's own head-related transfer function/head-related impulse response [151] [161] [162].

Wenzel *et al's* 1993 [151] study implemented a system HRTFs based upon that of an individual not included in the study group, finding that non-individualised head-related transfer function measurements led to increased front-to-back reversals, and degraded elevation localisation accuracy; likely a result of the spectral cues generated by the pinnae of the HRTFs not agreeing with those of the test participants. Moller *et al* [161] tested not only localisation accuracy with individualised and non-individualised measurements vs. real world localisation but also the hypothesis that individuals may adapt to non-individualised measurements to gain improved localisation ability with them.

Figure 4.g A participant positioned with test apparatus, and "digitizer" to indicate the perceived location of sound sources in the Experiments of Moller et al [102]

Eight test participants were placed in a room, one at a time, in the presence of 19 loudspeakers arranged at various angles and distances from the listening position. Binaural recordings of each speaker producing a stimulus sound (female speech) were made by placing microphones in the ear canals of the participants. The microphones were removed and the stimulus was played form each speaker in a random order, whilst the participant indicated the perceived location of the sound source via a “digitizer” tablet. Tests were repeated with the participant then listening to a random order playback of the stimuli recorded from their own head, that of one other test participant, and a mixture of the two conditions. They reported no significant difference in results between real-life and individualised recordings, with non-individualised recordings they found the highest concentration of errors on the median plane as front-to-back reversals or elevation misjudgements and a smaller, though still significant, number of distance errors. Although they found no significant adaptation of participants to the non-individualised recordings, it should be noted that their experiments did not include any feedback by which an individual could determine if they were making localisation errors or not- a factor which may have hindered any learning effect.

In their study of issues relating to binaural virtual reality, Begault *et al* [162] directly compared the effect of non-individualised HRTFs, acoustical reverberation, and head tracking on localisation accuracy, in-head-listening, and front-to-back reversals. It was found that individualised HRTFs did not significantly improve front-to-back reversals, in-head-listening or azimuthal localisation accuracy in a system where reverberation was included, although errors in elevation were still observed under these conditions with non-individualised HRTFs.

Given that making individualised head-related transfer function measurements is a time-consuming undertaking, even when interpolation is employed to reduce the total number of measurements required and has implications for scaling a virtual reality system to suit many users, the use of features such as simulated reverberation and head tracking have become central to the debate over creating effective virtual reality systems.

4.2.2 The Necessity of Head Tracking and Head-Movement Emulation

Head tracking is a common feature of current, commercially available, virtual reality systems [163] [164] [165] and allows the movements of the user's head to be translated into control data for head mounted video and audio display systems. In entertainment applications, where a common use virtual reality is visual-centric video games, head tracking has become a standard feature of hardware. In assistive auditory virtual reality, although head movement has been demonstrated to improve auditon in terms of front to back reversals[29] [149] [166] [150] [167] and in head localisation[30] [168] [5], the necessity of head tracking has been questioned [6]. Furthermore, the design criteria which are relevant to effective head tracking for audio based systems has been a long standing topic of study and debate [169] [3] [170].

---

[29] A localisation error where a sound located in front of the user is perceived as being located behind them, or vice versa

[30] A phenomenon where the illusory 3D sound environment delivered via headphones collapses and sounds are perceived as originating from within the user's head

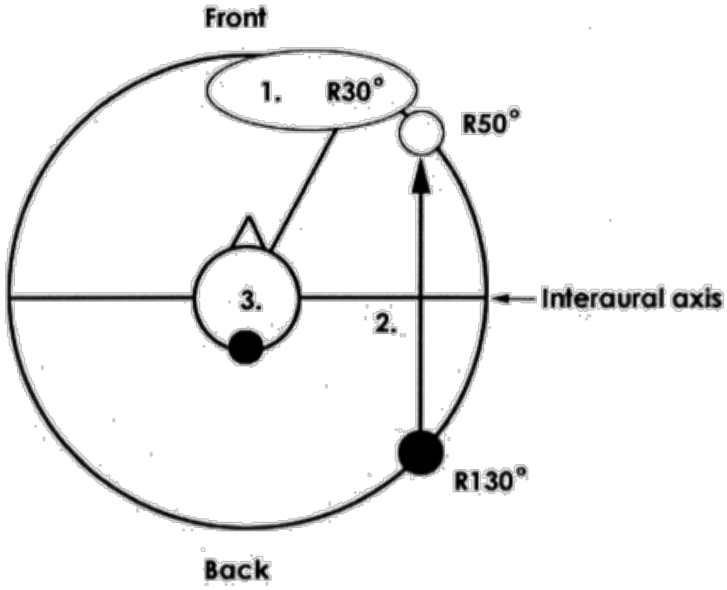


Figure 4.h Diagram illustrating typical errors in virtual acoustic headphone studies. 1.) localisation error: deviation in perceived source location compared to target location. 2.) Front to back reversal: stimuli perceived as originating on the opposite side of the interaural axis compared to the target location. 3.) In head localisation: distance error presenting as a source perceived inside the user's head rather than in the environment about them. (after Begault et al 2001)

In her paper calling for virtual reality systems developers to appropriately consider human perceptual factors Wenzel [149] noted that in her own studies the use of head tracking in auditory virtual reality had reduced the instance of front-to-back reversals from 28% (without head tracking) to 7% (with head tracking). This finding supported assertions previously made by Wallach [171] that small head movements would resolve localisation ambiguities arising because of what is now known as the cone of confusion (discussed in *2.6.4*). In general, head movement has repeatedly been indicated to improve sound localisation accuracy [172] [166] and so the argument of whether or not to implement the use of head tracking may seem to be resolved however, in a 1999 study by Wightman [150] it was demonstrated that user head movement was not necessarily required. In Wightman's study participants were presented with localisation tasks in which head movement was either encouraged or restricted; as expected, head movement reduced front to back reversal significantly, but in a second phase participant head movement was restricted whilst the location of the sound source itself was mobile and controlled either by the user, or by the researcher. When participants were able to control the source movement, front-to-back reverslas almost disappeared. This was an early idication that manual controls were useful in ameliorating localisation issues- in a virtual environment, sound sources moving consistently with one another in response to user input would create perceptual cues equivelant to movement of the head.

With the aim of investigating the necessity of head movement for navigation in auditory virtual reality, Blum *et al* [173] took the study of head movement from static localisation scenarios into a navigational task. Participants were placed into a virtual room with the task of navigating to target locations using auditory feedback alone. Tests were conducted on 20 participants in groups placed under one of two head movement conditions: joystick control for environment exploration with separately tracked head movement emulation, and joystick control with no emulated head movement. Evaluation of the impact of each condition focused on how many auditory targets participants could reach, and the time it took to reach each subsequent target presented. Although there was a large variance in success rates (from 13-100% successful target 'hits') no significant effect was found for the head tracking conditions; success instead depended upon the participants's experience with video games, and a significant learning effect (summarised in *figure 4.*1) was noted across all participants in both test conditions.

Learning effect					
Trial	2	3	4	5	6
Mean hit time (sec.)	22.1	22.4	20.2	19.0	17.0
Standard Deviation	12.5	12.6	11.5	9.8	9.9
% of hit targets	78%	82%	85%	81%	91%

Figure 4.i Performance as a function of trial sequence in Blum (2006) (after [2])

On the topic of in head localisation, Wersenyi conducted studies [168] [5] in which sound sources presented on a virtual 2D plane in front of the listener, and were randomly moved about their target location by small increments; 1-7º movements were induced in the virtual sound source which was spatialised with non-individualised HRTFs, and 50 participants were asked to evaluate whether they percieved the sound source as originating from in front of them, behind them, or inside their head. They were also asked to indicate when they became aware of the movement of the sound source, at which point the test was terminated. It was found that movements between 1-2º were sufficient to reduce in-head-localisation for 14 of the 50 participants (it is worth noting that 28 of the participants did not encounter in head localisation errors with or without the randomised movements), although front to back reversals were not significantly affected by random sound movements- a fact that supports Wightman's discovery that such errors were only significantly improved when the user had control of sound source movement [150].

Whilst it is clear from the work presented here that dynamic changes in localisation cues that come from the movement of the user are useful in reducing problems of both front-to-back reversal, and in head localisation- and that movements of the sound itself may also reduce inhead localisation- it is also apparent that the most imporatant factor in effectively achieveing such improvements is that of control. Whether via head tracking, or some other manual control, a reduction in the occurance of front-to-back reversals were found in situations where the user was able to control his or her presence and movements within the environment, and that head movement relative to body position may not be necessary for user's to successfully explore a virtual environment at all.

4.2.3 Implementation of Head Movement in Auditory Virtual Reality

Although the need for discrete head movements (separate from those movements associated with a user's displacement and orientation within a virtual scene) may be debatable, it is still necessary to implement emulated movement in a way that allows the user to explore an environment as they require. Both studies, and guidelines have been offered on the topic of achieving emulated movement- of chief concern in this area is asynchrony between the movements of the user or environment and the cues associated with locating objects/sound sources. Although visual stimuli associated with a sound source may mitigate asynchrony problems [7], and several studies include cross modal analyses between auditory and visual stimuli [174] [175] [176], in the context of virtual reality for the visually impaired such comparisons are problematic: it cannot be assumed that the users will sense visual stimuli, since people with no functional vision are amongst the visually impaired population.

In an important early study of auditory virtual reality, Sandvad [169] investigated three dynamic parameters which impact upon stimulus synchronicity;

- System latency: the time between the transduction of a user's control movement, to the time the consequential changes in scene are rendered
- Update rate: the rate at which changes in user control devices/movmeent trackers are read by the system
- Spatial resolution: the angular resolution of head related filters applied to the audio presented to the listener, which is analogous to the number of discrete localisation cue angles available

These parameters were investigated with the goal of determining optimal system performance in respect of rendering the auditory scene whilst avoiding asynchrony. Whilst comparing participants' ability to locate the source of pink noise in both real and virtual (head tracked) anechoice environments, Sandvad [169] determined that increased system latency had the largest negative impact on localisation; latency of greater than 96ms degraded performance in terms of both time taken, and accuracy. He also noted that despite update rates of 20Hz introducing audible temporal inconsistencies in the scene, localisation performance did not significantly degrade until the update rate was reduced to 10Hz or less. Spatial resolution also had a less significant impact on participant performance, with little difference in error for head-related transfer functions[31] resolutions of between 1° and 13°, in spite of the degredation in resolution being audible. These results were in contrast to an earlier study by Bronkhorst [177] who, using a spatial resolution of 5.6°[32] an update rate of 30Hz, and a maximum stimulus duration of 15

[31] It should be noted that, in this experiment, Sandvad used measurements of the individual participant's head made at ≈11° steps, using interpolation to complete the "missing" measurments.
[32] Measured, not interpolated

seconds (where stimulus was only stopped before 15 seconds if the participant indicated that they had located the sound source), had reported that system latency of up to 150ms had not disrupted localisation.

Wenzel performed two studies on the topic of system latency, one focusing on long duration stimuli of eight seconds [178], and one comparing long and short (three second) stimuli [3]- studying the effect of latency not only upon localisation error, but also upon front-to-back reversals, in head listening/distance perception.

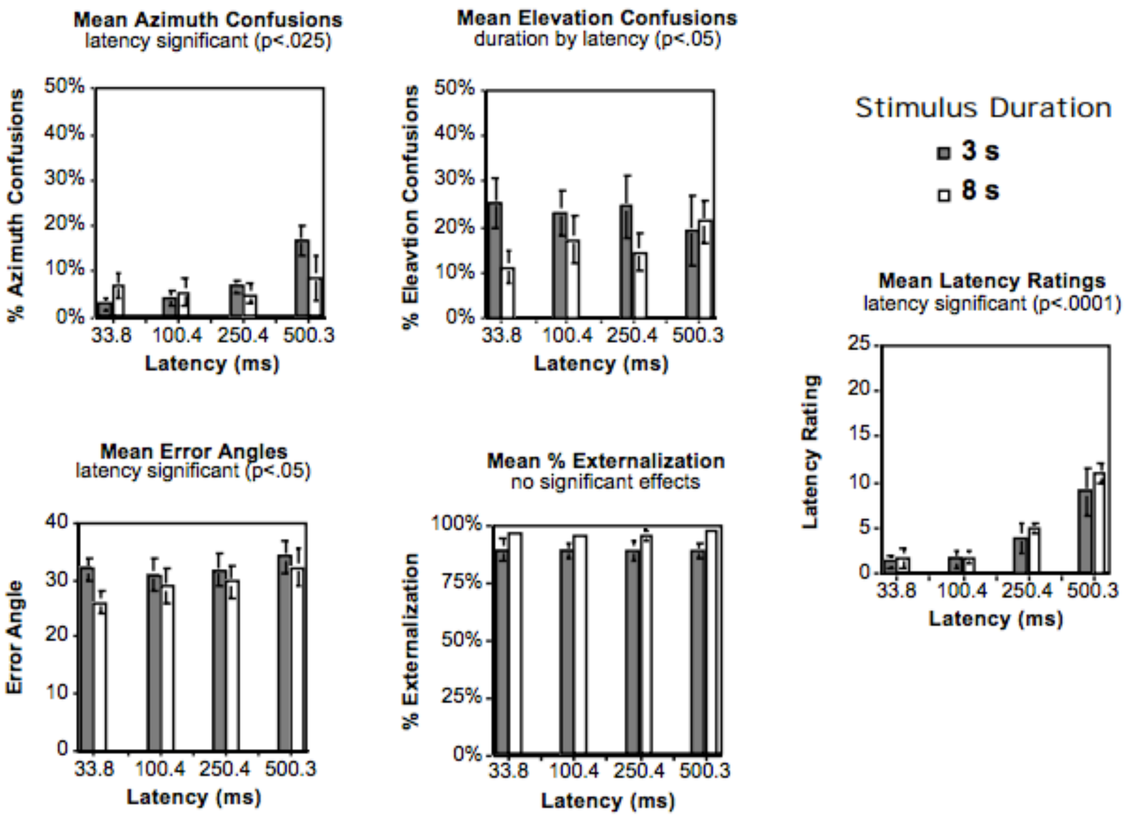


Figure 4.j The impact of system latency on auditory localisation; localisation errors (error angle), front to back confusion (azimuth and elevation confusion) and in head listening (externalisation) for 3 second and 8 second stimuli [4]

As *figure 4.j* demonstrates, aside from in head listening (given as externalization) the longer stimuli tended to be less affected by increaing latency, although there was only a significant difference for mean elevation confusions. Latency was a significant factor in both mean localisation error angles, and mean azimuthal front-to-back confusion at either stimulus length.

Brungart *et al* [179] investigated the interaction between system latency on stimuli of durations ranging from 64ms to 2s, the system had an optimal latency of ≈10ms with latencies from 0ms to 200ms added for the purpose of the experiment. In this experiment non-indivudiualised head-related transfer function measurements made to a 1º resolution using a KEMAR dummy head. Unlike Wenzel's study [3], Brungart *et al* measured the response times of participants and, in a second phase of the experiment, introduced a two second response time limit, forcing participants to log the percieved sound source location during this window. In terms of localisation error, the results were similar to those of Sandvad [169]; latencies of 73ms or less offered optimal localisation accuracy (in fact latenicy as low 12ms did not show a significant improvement over that of 73ms). In contrast to Wenzel's [3] study, latencies greater than 143ms elicited significant front-to-back confusion which Brungart *et al* note is likely a result of the restricted response time employed in their test.

The relationship between system latency and user movement was investigated from a different perspective by Suzuki *et al* [180] who looked at whether or not asynchrony casued by latency influenced participant's head movements; participants were asked to turn their heads to face sound sources presented with total latencies of between 12ms and 2s, when they had located the percieved source of the sound they indicated its direction by nodding. Latency of 500ms induced large "overshoots" in participant head movement, although reasonably accurate localisation was still achieved by the participant readjusting their head position, however the time taken to achieve localisation (with latencies of 500ms or more) increased proportionally at twice the value of the system latency. Although overshoot could be expected for perceptible latencies, as the location of the sound source would not be presented correctly during the initial onset of movement, it is interesting to note that localisation is still possible- albeit with some delay compared to localisation at lower latencies.

It is clear that latency poses a problem for auditory design, particularly in the context of assistive technology; individuals relying upon sound cues to locate target destinations in wayfinding, or avoid audible hazards in training simulation scenarios could potentially be misdirected. Further, the use of certain features to mitigate phenomenon such as front-to-back errors (such as integrated head tracking and body related impulse responses, emulated room acoustics etc.) may unfortunately degrade user experience and performance due to increased latency.

4.3 Virtual Room Acoustics

The virtual emulation of acoustical spaces is fundamental to creating a convincing and immersive virtual reality [181]. Sound reflections from objects in the environment have also been proven useful to the visually impaired when wayfinding and building mental maps of a space [19] and, as discussed in *2.1.8*, is generally useful to all humans in establishing sound source distance; the emulation of room acoustics is therefore of great interest to virtual reality system designers. Furthermore, reverberation has been shown to mitigate the issue of in-head localisation in binaural systems [162], and so it may have extra value for virtual reality designers.

This section presents studies in room acoustics for virtual reality, on topics ranging from the perceptual impact of introducing reverberation to a virtual environment, the level and fidelity of virtual sound reflections required to enhance localisation tasks, and the directional accuracy required when rendering sound reflections in virtual reality.

4.3.1 The Effect of Reverberation on Auditory Localisation

Although the precedence effect (discussed in *2.6.5*) might be expected to mitigate localisation problems caused by early reflections, there are instances in which the characteristic echo suppression of precedence effect may be reduced or broken down.

The Clifton Effect refers to a phenomenon first described by R.K Clifton [182], which is observed when a click train composed of a single click presented from the left or right, followed by click presented on the opposite side, with a time delay within the echo suppression threshold. If the sides at which the click and its delayed counterpart were suddenly swapped, echo suppression collapsed and each click was localised as a separate sound source; subsequent presentations of the clicks in their new positions led to echo suppression building up until the clicks were once again perceived as originating from a single location (that of the leading click). Blauert and Col [183] repeated the experiment whilst alternating the side that the leading and lagging click were presented on. The Clifton Effect was observed after each position change however, after several alternations listeners only perceived the leading sound as moving in position, indicating that echo suppression could “build up” as the listeners learned the position of reflector and source over numerous presentations. Yost and Guzman [184] showed that simple movements of the sound source and reflection were not enough to produce the Clifton Effect; lateral reversal of the positions was required to confound echo suppression.

Whilst reverberation has a clear impact on distance discrimination (via direct to reverberant energy ratios), early reflections alone may mitigate the problem of in-head-listening [162]. However, reverberant energy also has the potential to confound sound localisation in both real and virtual environments [185]. In a formative study of the effect of adding reverberation to a system, Begault [22] implemented a synthetic reverberation system based upon that of [186] consisting of an early reflection pattern, and a late reflection pattern; 66 individual early vectors were modelled as having arrived from the boundaries of an asymmetrical room (two of these early reflections were designed to replicate reflections from the floor of the room, and arrived ≈12ms before the onset of the other early reflections). The late reflection pattern consisted of exponentially decaying mixed white and weighted (1/frequency) noise. As may be expected, Begault found a significant reduction to in-head-listening amongst test participants however, the presence of reverberation degraded localisation in both azimuth and elevation by as much as 23º; attributed to the presence of lateral early reflections, which agreed with the work of Hartmann & Rackerd [187] who had previously studied the perceptual effects of reflector positions in a real-world environment. Shinn-Cunningham [188] made empirical head-related transfer function measurements in a reverberant environment, noting that reverberant energy reduced the spectral modifications in direct sound associated with the pinnae; a possible cause of increased elevation errors in the presence of sound reflections. Furthermore, the effect worsened with distance (as the direct energy decreased) and was more pronounced at the ear farthest from the sound source (as the acoustical shadow of the head also reduced direct sound energy).

Although it may be tempting to remove or modify/simplify reverberation in a virtual reality system for the sake of improved localisation and system efficiency, the apparent fluidity of echo suppression first described by Clifton [182] hints towards a further function of reverberation (specifically, early reflections) for human auditory perception; echolocation, a topic which warrants consideration and is discussed in the next sub-section.

4.3.2 Locating Sound Reflective Objects Using Reverberation

Until approximately 1950 it was thought that blind individuals possessed some extra sense which allowed them to detect the proximity of objects in the environment, termed "facial vision", it was thought to relate to some tactile sense of proximity to an object before physical contact had been made. Cotzin and Dallenbach [189] finally dispelled this notion when they placed "the final piece of the puzzle" of facial vision; facial proximity to a surface was not sufficient to detect its presence, sound reflected from the object alone however, *was* sufficient. In following studies, it was discovered that visually impaired individuals could discriminate the distance, width, and material of sound reflecting objects in a lateral ±90º arc about the azimuth by way of detecting frequency and loudness changes associated with short delayed reflections and is particularly acute if participants are permitted to choose their own aural stimulus (such as clicking or hissing made by the mouth) [190] [191] [192] [193].

The echolocative abilites of the visually impaired have proven viable in both virtual reality, and recorded audio. Schenkman and Nilsson [194] recorded binaural audio in both a reverberant and anechoic room, consisting of 5ms, 50ms, and 500ms noise bursts, with the use of a mannekin head in the presence and absence of a .5m aluminium disc was placed at distances of .5m to 5m from the mannekin. Sighted and visually impaired individuals were tasked with determining which recordings presented to them contained the reflecive disc. Performance was slightly better in the reverberant room, than the anechoic, and all participants showed the best performance with longer duration sounds and a distance of less than <2m from the reflector; two visually impaired individuals were noted to perform excpetionally well at greater distances, and a further phase of testing revealed that this was not due to chance. In general the performance of visually impaired participants was greater than that of their sighted counterparts, especially in the reverberant environment.

A similar series of three tests was conducted by Picinali *et al* [195] in which differing sound sources were binaurally recorded in reverberant environments. In the first test speech stimuli were used to assess the ability of blind and sighted individuals to locate a reflecting baffle placed at different angles around a mannekin head, the somewhat unclear results showed no significant differences between blind and sighted individuals, with age appearing to be the most significant factor. In the second test the baffle was placed in front of the mannekin head at distances between 5cm and 2m, with distance incrementing in 5cm steps, a detection threshold of 2m (the limit of the test distance) was found for visually impaired participants, whilst the threshold for sighted participants was ≈1m. In the final test, participants were asked to determine the size of rooms using a forced choice paradigm in which an impulsive click had been binaurally recorded; recordings were made in four rooms of varying size and shape, and participants were presented with four physical scale models of the rooms and asked to match the models to the stimulus heard. The visually impaired group showed greater accuracy in doing so, although there was a wide range of performance in the group.

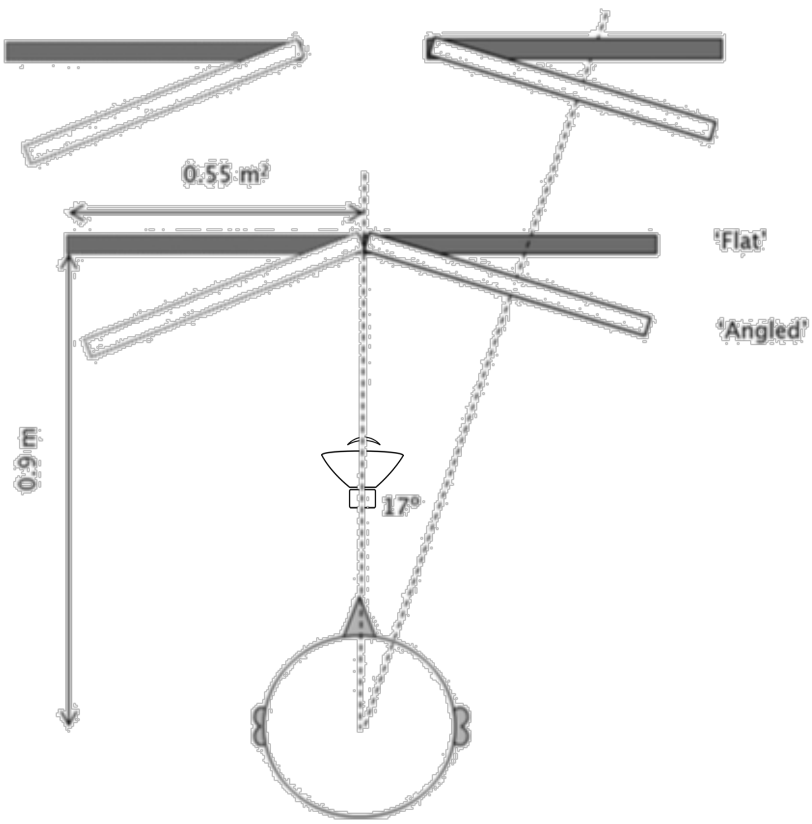


Figure 4.k Illustration of the arrangement used to study object echolocation in Rowan et al [129]. The speaker is positioned 25cm below the manikin head, and 5cm in front of it.

Rowan *et al* [196] conducted further studies of human echolocation by making impulse response measurements of a mannekin head in the presence of reflective boards placed alternately at the positions and angles shown in *figure 4.k.* Bands of synthesised, varying duration, filtered gaussian noise were convolved with the input responses to create stimulus of broadband (100Hz – 12kHz), low-pass (100Hz – 2kHz), and high pass (2kHz – 12kHz) natures. A single interval recognition[33] task was employed in which participants were presented with a single stimulus and a reflective board in one of the measured positions; they were required to indicate the perceived position of the board via button push, after which the correct position of the board was revealed to some listeners via an on screen message for 400ms. It was found that with boards positioned parallel to the interaural axis (in the "flat" position) echolocation for both sighted and blind

[33] For full details of the single interval recognition task see Macmillan & Creelman (2005)

individuals failed at distances >1.8m, whereas echolocation for certain memebrs of both groups extended to ranges of 3m for angled boards. This was attributed to the specular reflections available at both ears when the board was angled. Stimulus of >2kHz was significantly important to echolocation; where echolocation for flat panels required energy above 2kHz in the stimulus sound, and was primarily reliant upon it for angled panels. This would indicate that binaural cues are likely to be used for echolocating objects at larger distances. Interestingly, a phase of experimentation was conducted in which the direct sound emission was removed from the stimulus, so that only the echo could be heard, in this case listeners were still able to determine the direction of angled panels up to 3m away with 90% accuracy. In all experiments the sound pressure level was only found to produce weak differences in performance, most notable between 45dB and 55dB.

The work of Rowan *et al* indicates at the possibility that reverberant energy could be restricted in level to reduce the spectral notch “filling” discovered by Shin-Cunningham [188], with the possibility of mitigating certain sound source localisation problems associted with reverberant environments. Unfortunately such manipulations would also have the potential to confound auditory distance estimations as they would likely modify the direct to reverberant energy ratios. In any case the long documentary history of echolocation in visually impaired humans is a strong indication that accurate early reflections in an explorable environment can provide an effective method for identifying the presence of non-sound emitting objects.

4.4 Spatial Cognitive Mapping in Visually Impaired Individuals

The term cognitive map refers to a mental representation of a geographical location. It originates in Tolman’s 1948 paper *Cognitive Maps in Rats and Men* [197], in which the ability of rats to explore and learn mazes over repeated exposures was tested. In this work Tolman proposed that rats were not merely responding to external stimuli as they explored, but were in fact capable of generating cognitive maps as their exposure to a maze increased. The term mental map emerged later in literature and was popularised by Kevin Lynch in his work on urban planning, and human perception of built environments, most notably *The Image of the City* [198] which was concerned with studying human cognitive maps of cities via sketches created by individuals with some familiarity with the cities in question (namely Jersey City, Boston and Los Angeles).

Such maps are of great interest in the development of virtual environments for the blind, as numerous such technologies are concerned with assisting users in creating cognitive maps of real locations [18] [199] [19], and yet more studies of the efficacy of such equipment requires that test users demonstrate some aspect of a virtual auditory environment by sketching mental maps of it .

Several features of the visual system have been identified in demonstrating that it is the most suitable sensory modality for spatial representation [200] [201] [202] therefore, in the exclusion of this sense, the means by which visually impaired individuals form cognitive, or mental maps, and the efficacy of those maps has been a repeated area of study. Although early researchers such as von Senden [203] argued that spatial concepts were impossible in people who have been blind from birth (known as the Deficiency Theory), empirical evidence against this position has subsequently been accumulated. Consequently, two alternate theories have gained traction: the Inefficiency Theory, in which visual impairment leads to less efficient (rather than truly deficient) mental maps, and the Difference Theory, in which the formation of mental maps is qulitatively different, but not necessarily less efficient, for the visually impaired [204], whom may compensate for lack of sight in other ways [205].

The following section presents an overview of cognitive mapping and spatial understanding in the visually impaired, which is particularly relevant to designers aiming to achieve systems for "pre-learning" of existing environments, as well as those attempting to create simulations in which user exploration is a feature. It is common for map recreation in the visually impaired to be achieved by model building (using a simple but tactile medium such as construction blocks or LEGO®), directly recreating a particular route of interest by walking, or by estimating distances and directions (the latter often with the assistance of some pointing device). For comparative assessments of the effectiveness of these techniques see Haber *et al.* [206], and Kitchin & Jacobson [207].

These techniques have been used not only in exploring the cognitive maps of visually impaired people, but also as a means by which they can gain/retain information from technological systems designed to present maps and wayfinding information to them.

4.4.1 Evaluating Pre-Existing Cognitive Maps

In a 1978 study [208] of the cognitive mapping abilities of high school students, who were asked to reconstruct a model map of their school campus, Casey found that independent mobility correlated with the nature and accuracy of the reconstructions made by visually impaired students. Many of these students created maps featuring clusters of buildings connected by straight paths, when in reality paths may have been curved, and buildings somewhat distributed differently. Golledge *et al* [209] would later state that congenitally blind people showed greater accuracy in reconstructing features on routes which were more familiar to them, and that they tended to represent map features in compartmentalised segments, and to present curved paths as straight. This was not tested against a sighted group, although a partially sighted group did not show the same pronounced tendencies.

Further evaluation of the cognitive mapping abilities of younger visually impaired children (from ages 6 to 12) by Ungar *et al* [210] found that distance estimates of locations on their school campus correlated most closely with the functional distances between two points, rather than the euclidean- that is to say that distance estimates appeared to be based upon the routes which the children took when navigating the campus, rather than “as the crow flies” absolute distances form point-to-point. This led to children overestimating the distances from one point to another. Papadopoulos [211] conducted a particularly interesting study in which school children aged 15-16 were asked to construct a tactile map of the area around their school. The study involved sighted children presenting visually impaired children with a list of features, which the visually impaired children ranked in terms of importance to them in navigating. The visually impaired children were then asked to describe the area in question, prior to being given the tactile model. These descritptions were turned into textured maps, with each student correcting his or her map unitl they felt it accurately represented the environment. These maps were then compared for accuracy to the scale model produced by the sighted group.

It was found that independently mobile visually impaired children had a more complete understanding of environmental features, with those possessing some functional eyesight showing better performance than congenitally blind children.

In comparisons of visually impaired and sighted individuals, it has been demonstrated that under the correct circumstances, the visually impaired are capable of demonstrating cognitive mapping abilitites similar to those of the sighted. Sensory substitution has been shown to enable effective spatial awareness in cases where vision is replaced with haptic feedback [212], and effective cognitive mapping abilities when auditory and/or haptic feedback is available [213] [214].

These findings clearly demonstrate that the Deficiency Theory is at best an oversimplifaication; although blind and congenitally blind persons may not produce the detailed or accurate cognitive maps of their sighted counterparts, they do indeed posess spatial awareness and are capable of forming some level of cognitive map, albeit one which tends to feature distortions in terms of distance and relative position of features. This is understandable in light of Golledge *et al's* 1996 studies [215] [209]- any representation of space based upon one's own movement within it, and with limited sensory awareness of features not directly encountered, is necessarily limited.

4.4.2 Evaluating Cognitive Maps in Unfamiliar or Constructed Environments

Although insightful, studies focusing on the assessment of environments which are somewhat familiar to the participants present a potential deficiency which is apparent through their own results; since factors such as distance estimates are based upon an individual's own preferences for route navigation through an environment it is hard to absolutely quantify the extent to which cognitive maps may be affected and distorted by lack of visual feedback. For this reason studies conducted in unfamiliar, or purposely constructed environments, where individuals have no particular route preferences or preconceptions are an important facet of studying cognitive maps in the visually impiared. This section is limited to the discussion of real world constructed environments, although virtual environments have also been constructed for the same purposes (studies in virtual reality are treated in chapter 4.4.3).

In a comparative study of sighted individuals, the congenitally blind, and those with late onset blindness, Rieser, *et al* [216] [217] investigated participants' ability to identify their position relative to several "landmarks" in both real and imaginary conditions within a constructed test environment. Participants were blindfolded to eliminate differences in vision levels during the study and guided through the environment by the researchers from a start point to each of the pre-constructed landmarks in turn and were then asked to either imagine that they were standing at one of the landmarks and point in the direction of each of the other landmarks or were led in an obfuscating route to a landmark and asked to point to the others. In a final phase, participants were asked to imagine that they were at one of the positions they had been led to in the environment, and to identify the relative positions of the other landmarks.

It was found that sighted, and late onset blind participants could more quickly and accurately locate the landmarks from a real position in the environment than from an imagined one, whereas the congenitally blind participants showed similar performance in either condition, being more in-line with the (poorer) performance of the other participants in the imagined condition. The researchers suggested that in individuals with previous visual sensory life experience the perspective and relative directional structure of the environment was more readily understood, allowing them to update their position in a cognitive map more easily as they moved. The performance of congenitally blind participants leads to the conclusion that those individuals were essentially forced to imagine their position in either condition, having less general awareness of perspective and environmental structure. In a separate phase of the study, participants were led from a starting location to a landmark in the environment and asked to independently navigate directly back to their point of origin. Here it was also found that no significant differences between groups existed, however the complexity of the outbound path had the effect of an increased delay before participants began to navigate their return route; in cases where the outbound path included highly complex features such as doubling back to cross over itself the ability of all participants to navigate directly back to the starting location was noticeably worsened. Controversially, Loomis et al [218] would later replicate this experiment finding that no differences between groups existed, stating that they believed that the group selection method was a possible cause of this difference; in their study participants were selected based upon their ability to navigate independently whereas in Rieser, et al [216] [217] participants were found via agencies to support visually impaired individuals. Loomis postulated that users of such agencies may require greater mobility assistance and would therefore struggle to complete navigation tasks without assistance.

The study of visually impaired individuals navigational abilities should consider the complex of abilities of the members of this group, particularly in the context of developing assistive technology, although it does provide a point of caution for making assumptions that previous visual experience is a strong pre-requisite of accurate cognitive map forming. Despite the failure to recreate group performance differences in Loomis et al [218], it was still the case that the more complex of a path participants were directed along, the longer it took for them to propose an answer to questions regarding the environment and the position of landmarks, or themselves, within it. This would tend to indicate that the internal metric used for cognitive mapping without sight is based upon some internal measure of an individual's direct experience of that space, a hypothesis also proposed by Golledge et al [215].

4.4.3 Cognitive Maps and Virtual Reality Systems for the Visually Impaired

In a classic study by Thorndyke & Hayes-Roth [219] sighted participants were introduced to a real world environment either by personal exploration, or by studying a map drawing. Map learners estimated functional route distances and euclidean distances equally well, however those who explored the environement made far more accurate estimates of route distances. It was concluded that environmental exploration led to a procedural knowledge of routes and locations within an environment. Following this study, interest in using virtual environments to create cognitive maps grew. Several studies were conducted comparing the effectiveness of virtual environments to that of maps, or real world training for spatial representations and cognitive maps. Many such studies indicated that training, whilst useful, was inferior to map or real world training [220] whilst others have demonstrated the efficacy of learning and navigation soley in virtual environments [221] [222], particularly after multiple training phases [223]. Following these demonstrations of the general usefullness of virtual environments, and the work discussed in *4.4.2,* interest in studying virtual reality environments to assist visually impaired persons in developing cognitive maps continued.

Afonso *et al* [224] implemented an augmented reality system in which sound sources, obstacle proximity alerts, and reverberation were rendered binaurally in a virtual representation of a room. The experiment took place in the real world counterpart of the virtual room, allowing users to navigate the environment by walking whilst the virtual scene was updated via wearable trakcing devices. Although the latency of the rendered audio was notably high (measured at a maximum of ≈250ms) it was found that, after a learning phase, distance estimation to targets within the envrionment was improved by virtual exploration of the environment compared to verbal descriptions of the environment.

In their 2014 study of systems and cognitive mapping in the visually impaired, Picinali *et al* [19] compared pre-recorded binaural and ambisonic presentations of an environment, an interactive virtual representaion, and real world exploration of the same space and assessed individual's ability to recreate the space as a LEGO® block map. It was found that passive exploration via recordings was the least effective method, whilst virtual and real world navigations of the environment produced comparably favourable results. The authors highlighted two features of their system which were significant in producing these results; a high level of interactivity/mobility through user control (in this case a joystick) and the ability for the user to "self generate" simulated noises within the environment such as footsteps, or finger clicks, as a sound source for echolocation. In an earlier review of virtual reality environments for the visually impaired, Lahav [18] noted that of 21 iterations of virtual reality systems reviewed only ≈20% included simulated footstep sounds. The absence of these sounds in systems geared towards user exploration is interesting, as not only do they form a potential sound source for user echolocation, but may also provide some displacement cue, indicating to the user their approximate speed and distance of movement; although the latter is apparently untested in literature so far, the inclusion of footstep sounds as a possible dual purpose exploration cue is appealing. It is also worth noting that in the same review, Lahav noted that virtual reality systems which required the user to operate multiple devices (such as head trackers and joysticks or haptic feedback devices) generated extra cognitive loads which may impact upon the user's ability to gather and analyse information about the environment leading her to call for system simplification to be borne in mind in future system designs.

4.5 On the State of the Art in Auditory Localisation and Virtual Reality

The perceptual boundaries of human audition and sound localisation have been a topic of longstanding interest to psychologists and technologists alike. Advances in the understanding of the duplex theory of sound localisation have led to the development of technologies designed to exploit this understanding and create new sensory human-computer interfaces based upon it. Work such as that of Loomis *et al* [218], Golledge *et al* [209], Darken & Sibert [221], Ruddle *et al* [223] and more has given a strong indication that mental maps can be formed via non-visual experience, strengthening the case for assistive virtual reality for the visually impaired; a system allowing users to experience simulations of environments and scenarios in safety could certainly aid them in attaining environmental understanding.

Whilst a significant body of work in this field demonstrates that spatialized three-dimensional sound can be achieved over headphones via binaural rendering, issues of localisation accuracy with the potential to undermine auditory virtual reality as an assistive technology remain. Begault [162], Shin-Cunningham [99], and Wenzel [178] [151] have made contributions to exploring and quantifying these issues which arise because of poor agreement between non-individualised impulse response measurements (HRIRs, and the associated frequency domain HRTFs, for example) and those of a given virtual reality technology user. Bronkhorst presented a system in which head-related transfer function measurements made at 5.6° intervals were complemented by interpolated head-related transfer function measurements, with no loss of localisation accuracy in the horizontal plane, although this made no indication towards solving the problem of the original measured HRTFs required agreement with those of the listener.

Features such as simulated reverberation and tracked head movement have been shown to alleviate the various problems of localisation in non-individualised measurements, except for perhaps elevation errors; Begault [83] showed that whilst reverberation greatly reduced in-head-listening it also increased azimuthal localisation errors, Begault and Wenzel [162] found that head tracking reduced front-to-back reversals and Thurlow [172], and Perrett [166] noted that it also improved localisation accuracy. Furthermore, Begault and Wenzel [162] demonstrated that a system which includes head tracking and reverberation, individualised head-related transfer function measurements were not significantly important. Although, Wersenyi [168] indicated that sound source movement below the perception threshold may be enough to partially reduce in-head-listening. Blum and Katz [173], and Katz and Picinali [8] conducted studies which demonstrated that for systems which include user displacement (free movement within the virtual reality environment) separate head tracking did not provide any clear benefits in terms of locating and navigating to sound sources when compared to manual head movement control via joystick. Reverberation also has secondary uses, particularly for the visually impaired, who can detect non-sound emitting objects via early reflections.

Although a compelling case for including as many of the above features as possible in a virtual reality system could be made, Begault and Wenzel [9] [3], Brungart [179], and Sandvad [169] have shown that excessive computational load can also introduce significant problems, making even more relevant Begault and Wenzel's earlier calls for designers to consider what is most necessary to avoid implementing unnecessary features in a virtual reality system. When considering virtual reality as assistive technology simpler systems may also be desirable from the point of view of cost effectiveness.

4.5.1 The Direction of Study Arising from the State of the Art

It is clear that there are potential use cases for auditory virtual reality as an assistive technology for the visually impaired- the study of cognitive mapping shows that experience of an environment can be helpful in establishing such maps and that the auditory and/or tactile sensory channels are useful for this in the absence of eyesight [212] [213] [214]. Visually impiared individuals have also shown an ability to develop cognitive maps form virtual experiences of an environment [18] [199] [19].

In spite of its potential, there are still issues of implementation in binaural virtual reality: it is possible for fundamental errors of sound localisation such as front-to-back reversal and in-head localisation to occur [150] [4] [151]. It has been demonstrated that there are a number of possible solutions to these problems, one solution is to use individualised HRTF measurements for the purpose of binaural rendering [151], although this method requires controlled measurement of each intended system user in an anechoic environment [146]. Another solution to these issues is to implement head movement (either tracked/user controlled, or emulated) into the system [149] [166] [150] [167] [168] [5], although in the case of manually controlled or tracked head movement this would instroduce the need for some control input. A solution to the issue of in-head localisation could be to introduce simulated reverberation and acoustics to the system [162] [22], although this increases the computational load of the system and may introduce latency issues [178] [3].

Given that the above issues associated with head movement and reverberation in virtual reality are issues of implementation which are directly tied to the system design itself, rather than to the system and to the user intended to work with it (as is the case for individualised HRTF measurements) it seems to follow that they could be solved by the designer and design process alone with no need for bespoke tailoring of the system to each user. Furthermore, it is known that as well as solving virtual localisation issues, reverberation plays a role in auditory distance estimation [84] [86] and an individual's ability to control movement through, or interaction with, an environment contributes to their ability to form cognitive maps of it [19] [208] [209] [212] [213] [214]. Therefore, it seems reasonable to include head movement and reverberation in an assistive virtual reality system.

The implementation of head movement does still have outstanding issues, studies using head tracking or movement typically guide the user towards this feature and whilst there is good indication that the sighted use head movement in virtual reality [225], there is little in the literature to substantiate the notion that the visually impaired would naturally do the same [2] [8] [134]. Indeed, there is some suggestion that the visually impaired, particularly those with early-onset sight-loss may avoid the use of head movement in some circumstances relevant to navigation [14]. This issue should be considered at the design stage, as users may require instruction to use head movement (either tracked or manually controlled) to facilitate audition and reduce the system deficiencies such as front-to-back localisation, and some such movements may in fact run contrary to natural behaviour. Furthermore, the implementation of reverberation into a system with head tracking poses a question to the impact that head movement/angle may have upon perceived reverberant cues and how this may be quantified in a way that is applicable to virtual reality designers.

To investigate these questions, a real-world field study was designed and is discussed in the following chapter. After this study concluded two further studies, a self-report questionnaire, and an acoustical analysis of mannequin HRIRs with body related impulse responses under the condition of head movement in a reverberant environment. These studies are organised into chapters: *chapter 5* details the creation of a measurement and analysis toolkit for the exploration study whilst *6, 7 & 8* detail the methods and results of the three studies.

Part II: Field Study

5. Analysis Toolkit

This chapter details the hardware and software used in the field study of head movement in "real world" urban navigation. The methodology and design of the study is detailed in *chapter 6.* This chapter is focused on the performance requirements for the hardware in question, and how they were met.

5.1 Hardware

The hardware used in the field study is comprised of a set of inertial sensors for motion tracking, an array of 4 microphones for audio capture, and a digital camera with 360° lens.

5.1.1 Inertial Sensor Package

The heart of the analysis tool-kit is the motion sensor package, which was used to capture test participants' head movements during their navigation of a semi controlled, urban environment.

The primary design requirements for the inertial sensor package were as follows:

1. The package must offer simultaneous tracking of head movements in 3 dimensions; yaw (horizontal plane head turning), roll (frontal plane head turning), and pitch (median plane head turning).
2. The package must be portable; it must have low enough power requirements to operate from a portable battery pack, it must be small and light enough to be head mounted without impeding head movement, and all tracking apparatus must be "on board[34]".
3. The package must operate in unpredictable environmental conditions; fluctuating electromagnetic fields[35], the presence of ferrous metals, and physical shock

[34] External tracking apparatus such as infrared cameras, or magnetic sensors would be impossible to deploy in the chosen test environment.
[35] Such as those produced by mobile phone antennae.

should not cause the package to fail to produce data, or to produce unduly erroneous data.

Considering these requirements, the Arduino compatible Razor[36] 9 Degrees of Freedom inertial sensor package was selected for use in the tool-kit. This sensor package is comprised of 3 x 3-dimensional inertial motion sensors (namely accelerometers, gyroscope, and magnetometer), the combining of multiple sensor types is common in the design of inertial sensors due to inherent deficiencies specific to different types of inertial sensor:

Sensor Type	Operating Principle	Common Deficiency
Gyroscope	Coriolis force [226]	Prone to drift: accumulates errors in reported position/movement slowly over time
Accelerometer	Conservation of linear momentum [227]	Prone to noise: reports small, rapidly fluctuating movements even when at rest
Magnetometer	Magnetoresistance [228]	Prone to environmental influence: reports momentary errors in position/movement in the presence of ferrous metals or electromagnetic fields

Figure 5.a The operating principles of common electronic movement sensors, their operating principles, and deficiencies

The varying operating principles, and the differences in potential errors/causes of error allow the data from the three types of sensor to be corrected by means of filtering (discussed further in section *5.3.1*).

[36] Model: Razor SEN IMU-10736 [270]

5.1.2 Microphone Array

In addition to measuring head movement the tool-kit also recorded the environment near the test participant. Audio recording was achieved via an array of microphones capable of recording sound in 360° around the participant.

As with the motion sensing package, there were several design requirements which needed to be met by the microphone array:

1. The array must offer recording of the local sound field, in 360° about the test participant (including vertical sensitivity).
2. The array must offer broadband frequency sensitivity, ideally of approximately equal bandwidth to the human auditory system
3. The array must be portable; it must have low enough power requirements to be battery operated, it must be small and light enough to be head mounted without impeding head movement.

To meet these requirements, an array of four omni-directional[37] electret condenser microphones were chosen. A summary of the specifications of each of these microphones is as follows:

Size	Weight	Power Requirement	Sound Pressure Sensitivity	Signal to Noise Ratio	Frequency Response
6x6x5.2mm	0.33g	1.5v	-60dBv	>40dB	20Hz-20kHz

Figure 5.b The operating parameters of the microphones used in the sensor array

[37] Of approximately equal sound sensitivity in all dimensions around the capsule.

The microphones were arranged equidistantly around the circumference of the elasticated headband of a hat, such that the wearers head would act as an acoustical impedance between the microphones. This would allow for differences in pressure and arrival time to be used to determine the approximate direction of a sound source in post experiment analysis, where necessary.

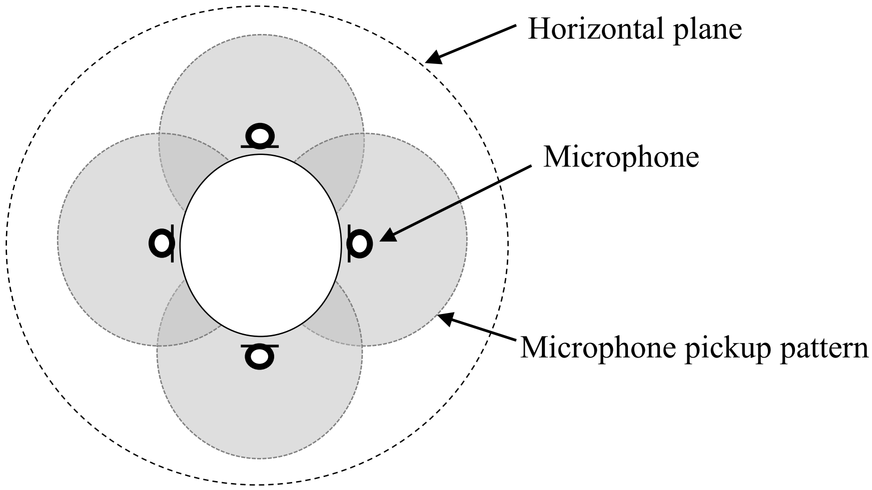


Figure 5.c Diagram showing the position of microphones relative to the test participant's head, with sound pickup pattern overlaid

The microphones were powered by a 9v battery pack and split rail power supply, worn in a small backpack carried by the test participant. Although fine localisation of sound sources was not expected of the microphone array, it was included in the sensor package to give coarse information on the local sound field as a potential aid to analyse patterns in head movement if no location or event-based pattern was apparent.

5.1.3 360° Camera

Due to the coarse nature of sound localisation available via the microphone array, and an anticipated need to verify the test participant's navigational route through the test area was correct, a camera was added to the tool-kit to provide visual data regarding the participant's immediate environment.

The design requirements for the camera were as follows:

1. The camera must offer full 360° video recording about the horizontal plane of the test participant.
2. The camera must be portable; have low enough power requirements to be battery operated, and it must be small and light enough to be head mounted, without impeding head movement.
3. The camera must have sufficient frame rate and image quality for footage to be intelligible.

To meet this specification, a commercially available five mega pixel digital webcam was selected. Its small size and USB connectivity made it ideal for use in the toolkit. Additionally, the camera had a simple manual focus control in the form of a ring mechanism around the lens, this allowed for focus to be set and fixed prior to each test.

Figure 5.d The 360-degree lens used to modify the digital camera for observational studies (left) and a frame of unprocessed footage taken from it (right)

To offer 360º recording about the test participant, the camera was modified with the addition of a mirrored, domed lens. This allowed the camera to be positioned on top of the participant's head, angled directly upwards, meaning that its field of vision was not occluded by the other components in the toolkit.

Although the footage captured was of a less than optimal quality (due to the extremely short focal length required to film through the reflective lens) the footage was adequate for its intended purpose in assisting with the identification of any ambiguous sounds, and their point of origin.

The final assembly of the hardware consisted of the head mounted sensor package (with inertial sensors, microphones, and camera) weighing ≈190g, and a backpack containing a laptop and power supply for the microphones, weighing ≈2.7kg.

5.1.4 Implementation of Hardware

Data from the apparatus was recorded to a laptop held in a small backpack carried by test participants (data recording is described in section *5.3)*. The inertial sensor was connected via Bluetooth, while the camera was connected via USB, and the microphone output was converted to a digital signal via and audio interface, which in turn was connected to the laptop via USB.

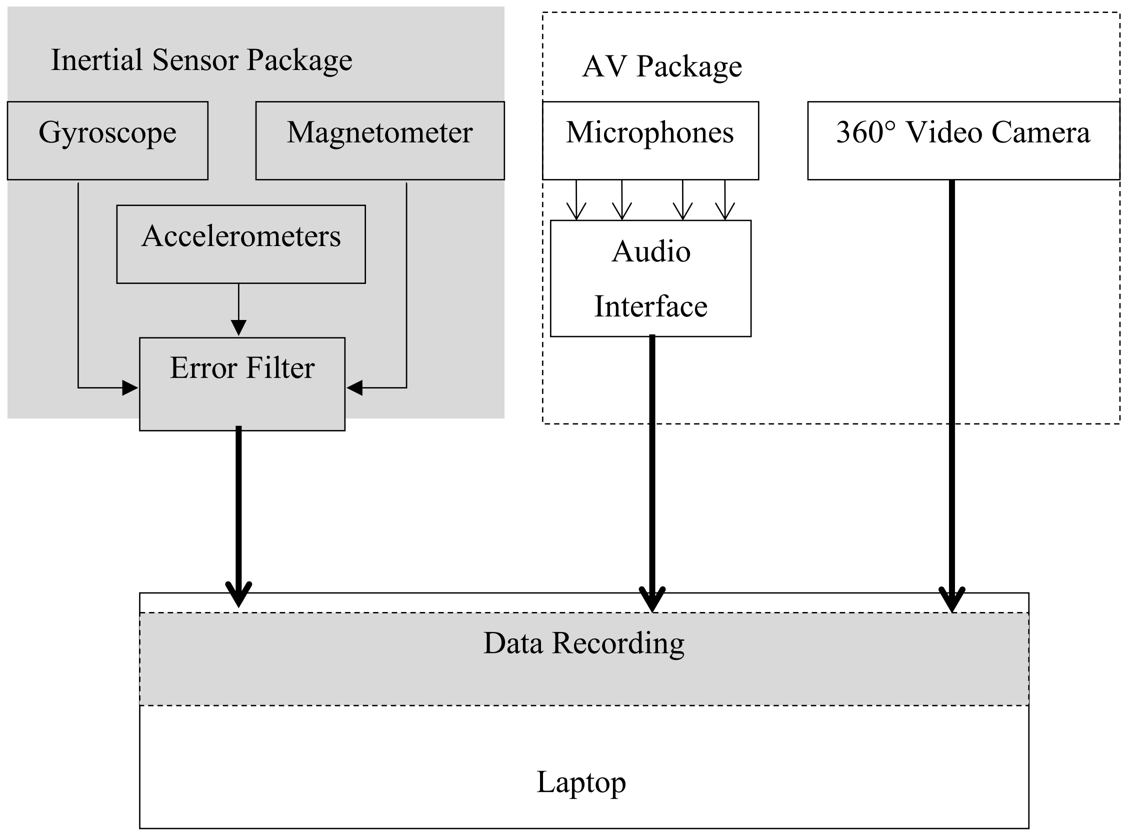


Figure 5.e Simplified signal and data flow diagram showing the connection of sensors, microphones, and camera to the recording device used in the analysis toolkit

Aside from the error filter implemented in the inertial sensor package's firmware, and basic sensor data formatting, all data integration and analysis was carried out post-recording (described in section *5.4*).

5.2 Data Capture Software

A novel system for data capture was created, to allow for synchronisation audio, image, and inertial sensor recordings; essential for proper comparison of participant behaviour in response to environmental factors such as the local sound field during the real-world test phase. The following sections describe the recording and synchronisation software, as well as the formats in which data was recorded.

5.2.1 Inertial Sensor Data Capture

The inertial sensor package utilised firmware and inertial sensor error filtering (using a Kalman filter model [229]) originally designed/implemented by Peter Bartz at the Deutsche Telekom Laboratories [230]. This allowed the unit to output sensor measurements as raw binary data via a connected Bluetooth shield. Further modifications were made to integrate video and audio recording/control for the sensor package, and to record inertial sensor data for later analysis (the final version of this firmware is available in *Appendix One*).

The Kalman filter assumes that the true state of a system is "hidden" by noise and/or errors generated by the system itself. In the application used in this work, it uses error models which are based upon observed control measurements of sensor errors in the inertial measurement unit (IMU) mixed with Gaussian noise. The Kalman filter assumes that the true state of the IMU at time *k* is evolved from the state at $k-1$ according to:

$$\mathbb{x}_k = \mathbb{F}_k \mathbb{x}_{k-1} + \mathbb{B}_k \mathbb{u}_k + \mathbb{w}_k$$

Where $\mathbb{F}_k$ is a state transition model, which is applied to the previous state $\mathbb{x}_{k-1}$, $\mathbb{B}_k$ is the control input model applied to the control vector $\mathbb{u}_k$, $\mathbb{w}_k$ is the process noise which is drawn from a multivariate Gaussian distribution.

At time *k,* an observation (or measurement) $\mathbb{z}_k$ of the true state $\mathbb{x}_k$ is made according to:

$$\mathbb{z}_k = \mathbb{H}_k \mathbb{x}_k + \mathbb{v}_k$$

Where $\mathbb{H}_k$ is the observation model which maps the true state into the observed state and $\mathbb{v}_k$ is the observation noise. For full details of the implementation in this work see Bartz [230].

The sensor package was controlled via Bluetooth using a Pure Data (PD) sketch, also designed by Peter Bartz [230]. This allowed the binary data derived from the sensor package to be formatted as a plain text file, with a single frame of heading data occupying three lines:

```
yaw 0.0
pitch 0.0
roll 0.0
```

The “zero bearings” shown indicate that the sensor has not moved from the position it was in when initialised.

The constant stream of output that was provided, once the sensor package was connected to the laptop via Bluetooth, proved to be impossible to synchronise with data recorded from the camera and microphones. For this reason, the Pure Data sketch was modified to include a time stamp (in milliseconds) at the start of each frame of data:

```
time 0
yaw 0.0
pitch 0.0
roll 0.0
```

This allowed for the “up time” of the sensor package to be determined, as well as the specific period over which any head movements had occurred.

Since the camera was also connected to the laptop (via USB), the PURE DATA sketch was further modified to handle camera output recording, such that when a predefined key on the laptop QWERTY keyboard was pressed, the sensor data was initialised to 0 and the video recording began, as well as a line of text being added to the sensor package output file to indicate when initialisation had occurred:

```
*****/////Alignment initialised, video started /////*****
time 249
yaw 0.0
pitch 14.70869
roll -12.779698
```

This allowed for inertial sensor data and video footage to easily be synchronised in post-test analysis.

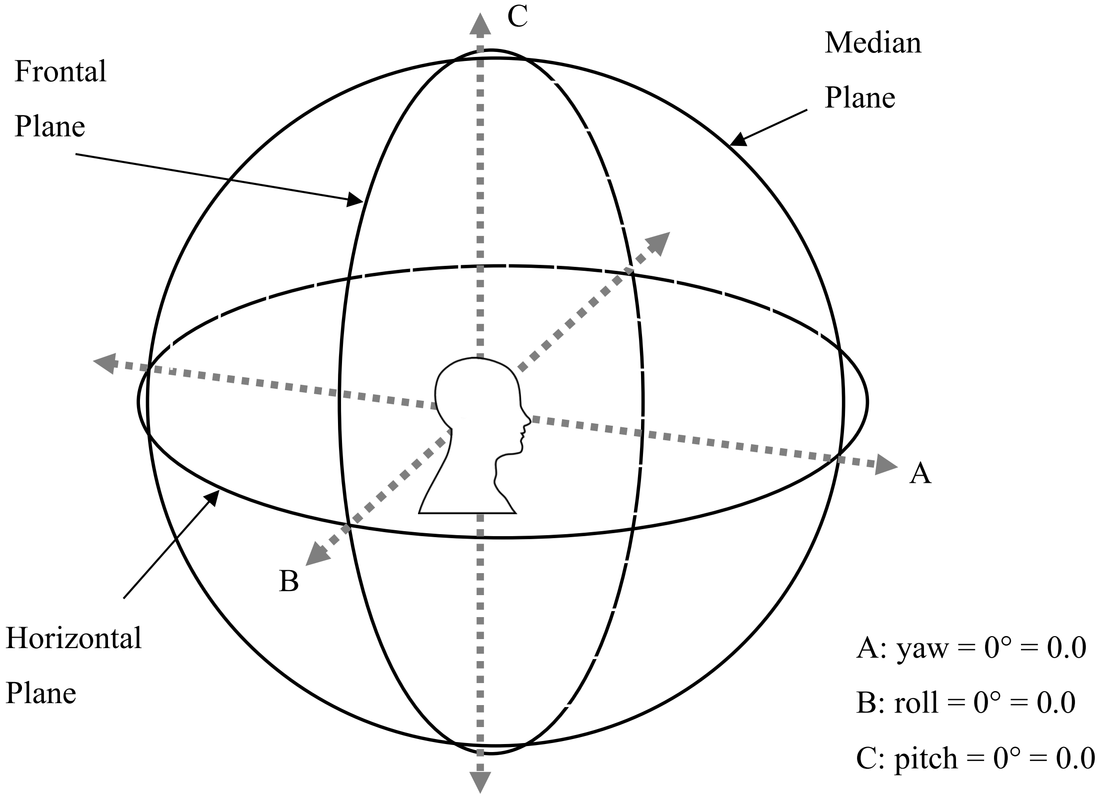


Figure 5.f Diagram showing the orientation of the test participant's head relative to the sensors reported orientation. At initialisation the inertial sensors default "zero bearings" are locked to the current position of the participant's head.

At initialisation, the inertial sensor package was calibrated such that its current attitude became the origin of any bearing measurements made. This meant that a participant's resting head position would become the reference point from which all subsequent movements were measured.

Movements of the inertial sensor were represented by an attitude heading reference system (AHRS) in which turns were represented by the following changes in reported data values:

- Yaw +ve = rightward azimuth rotation in the horizontal plane
- Pitch +ve = upward azimuth rotation in the median plane
- Roll +ve = leftward azimuth rotation in the frontal plane

5.2.2 Audio Synchronisation

Finally, the pure data sketch was modified such that a test tone of 1kHz would be played for 1 second, coincident with the alignment initialisation and the start of video data capture. This allowed for an audible indication of the test start conditions to be captured via the microphones, whose output was recorded as .wav audio (at 16bit/44.1kHz[38] by Audacity[39]). The final version of the pure data sketch is available in *Appendix One.*

5.3 Analysis Software

The focus for analysing data recorded during this experiment was twofold: to quantify any differences in head movement behaviour between sighted and visually impaired participants, and to identify any trends or correlations between head movement and environmental events/locations during each participant's test. MATLAB functions were created to analyse the head movement data captured by the inertial measurement unit during each test. The details of each analysis are described throughout *chapter 6*. This section serves as an introduction to the principles applied when designing appropriate analysis methods.

5.3.1 Isolating Data of Interest

Due to the test conditions requiring that participants walk whilst under measurement the analysis of head movement data was confounded by the capture of so-called head perturbations[40] in addition to the data of interest.

To reduce the impact of head perturbations upon result data it was necessary to filter the data captured by the inertial measurement unit. Under experimental conditions the mean maximum head turn (yaw) frequency for human participants is 2.6Hz although some extraneous head movement may occur at frequencies as low as ≈1.5Hz [231], a low pass filter for head movement data of 3Hz would eliminate all but the lowest frequencies of unwanted head yaw oscillation without interfering with head movements in the lower bounds, or the measure of an unmoving head (such as suggested for posture control [14].

---

[38] Standard compact disc quality digital audio [276]
[39] A free audio recording software
[40] Unintentional head movements arising because of body motion/gait during locomotion

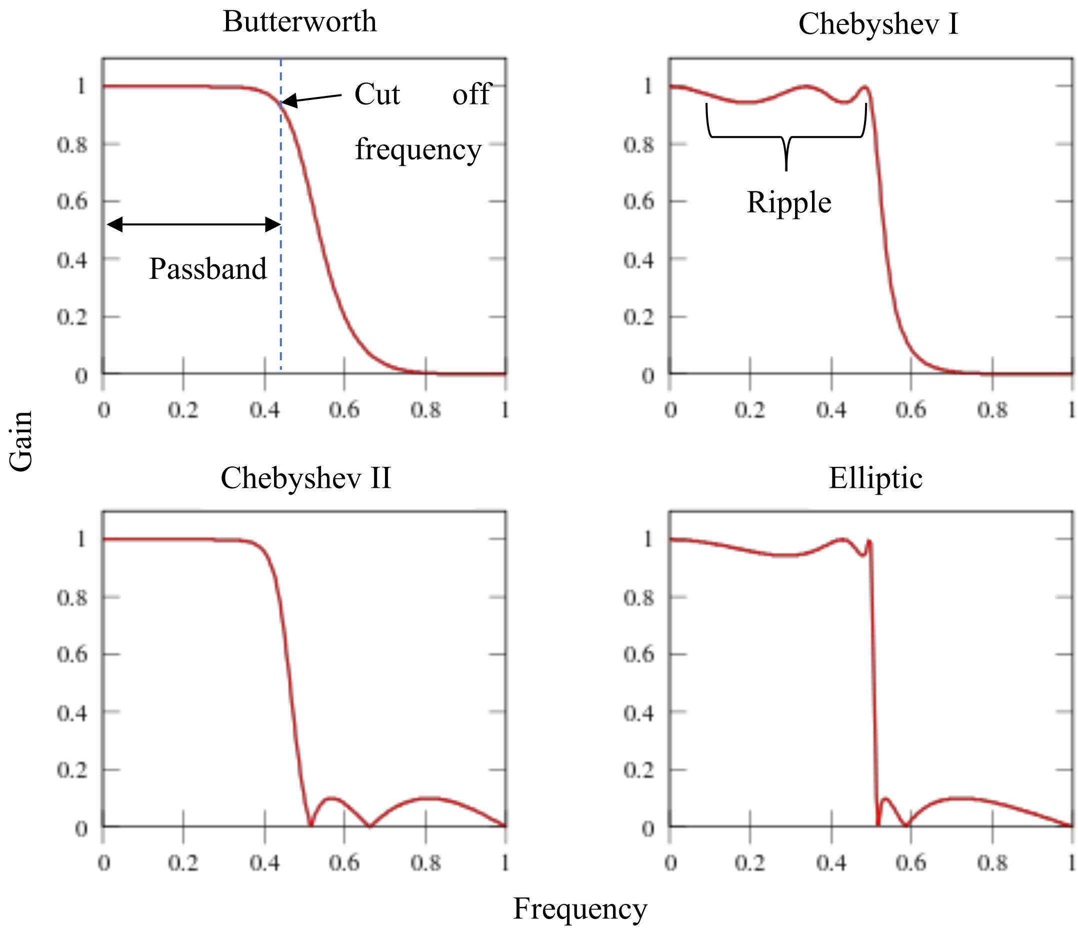


Figure 5.g Example plots comparing gain over frequency of four common filter types: Butterworth, Chebyshev (I & II) and Elliptical, after [219] with cut off frequency, passband and ripple illustrated

An ideal filter type would be one that introduces minimal ripple[41] to data within the passband and stopband whilst offering a maximal roll off (gain reduction over frequency) above the cut off frequency. For these reasons, a Butterworth type filter was chosen.

The frequency response of a Butterworth filter is given as:

$$H_{(j\omega)} = \frac{1}{\sqrt{1 + \varepsilon^2 \left(\frac{\omega}{\omega_p}\right)^{2n}}}$$

[41] Magnitude distortions which vary over frequency

Where n represents the filter order, ω is equal to $2\pi f$ (f = frequency over period) and ε is the maximum gain in the pass band. By increasing the filter order *n,* it is possible to increase the rate of roll off in the filter, rejecting frequencies closer to cut off without creating ripple (shown in *figure 5.h*).

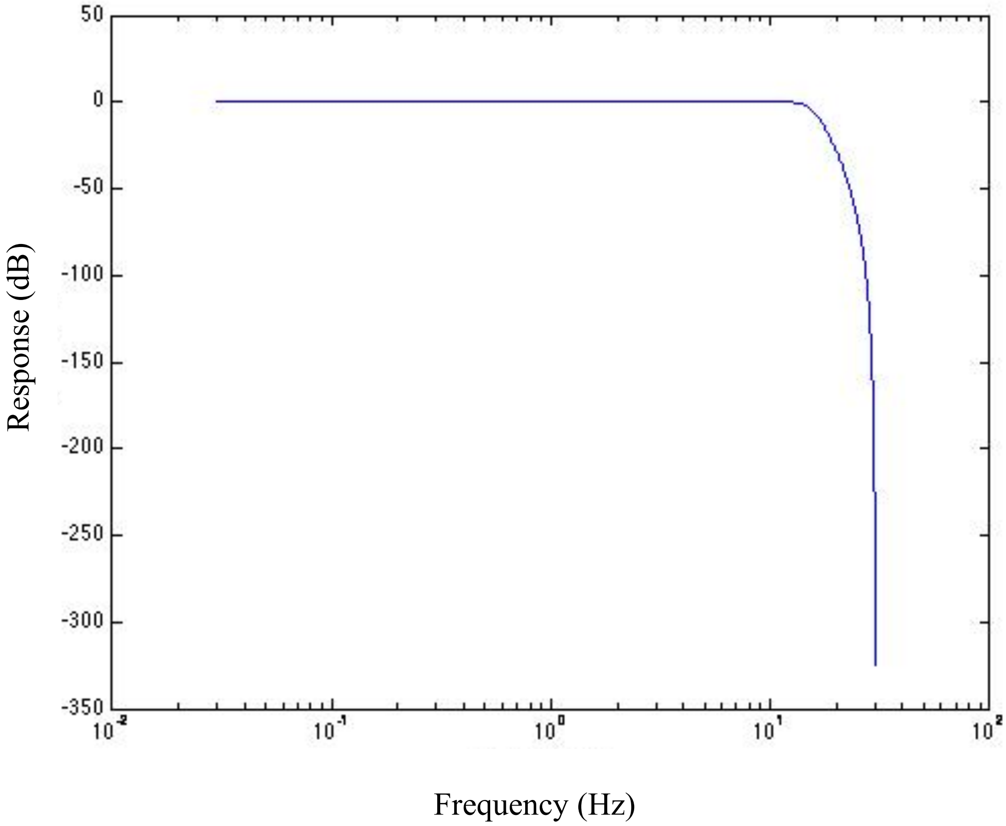


Figure 5.h Plot showing the frequency response of a 6th order Butterworth filter designed in MATLAB

For full details of Butterworth filter information see MATLAB [232].

6. Head Movement Field Study

This section details the methods and results of a small-sample field study of the use of head movement in navigation and auditory scene analysis. This study was conducted using the tool-kit described in *sections 5.1-5.3*.

The goal of this study was to compare the behaviours of visually impaired individuals with those of sighted individuals during the navigation of a prescribed route in an urban environment. Although the route and environment were chosen for the auditory stimulus likely to be present during participant navigation, no extra measures of control were taken in respect of the environment or sound field encountered by participants.

The study is considered limited as only four visually impaired test participants were available to participate. It was however, considered potentially informative, as there is currently no known data regarding differences, or lack thereof, in head movement behaviour observed in blind and sighted individuals under the conditions of this study.

The measurements obtained during the study were analysed first by producing descriptive statistics (the mean frequency of head turns per participant under the various conditions of the test route), followed by a series of unpaired t-tests (with 95% confidence intervals) to compare the head movement behaviour of sighted and visually impaired groups throughout the test, and then at specific stages of the test route. The various comparisons and analyses begin in *section 6.3*.

6.1 Rationale and Hypotheses

6.1.1 Rationale

The field study was motivated by a need to address the question of whether the visually impaired actively use head movement in navigation and auditory perception. Previous studies have demonstrated that sound localisation accuracy, for example, may be affected by the angle at which a sound is presented relative to the head. As such, either movement of the head or of the sound source may facilitate better localisation. Some studies have even suggested that their participants use head movement to assist in the completion of various audition-based tasks but, whilst these studies offer a compelling case for the fact that head movement can be successfully used to improve aspects of auditory perception, they do not necessarily show whether head movement is used naturally (without instruction) by humans in this way. Although the field study has its own issues which may affect ecological validity, such as the use of only a single environment and the need for participants to wear or carry measurement devices, the instructions to participants (detailed in *section 6.2.3*) intentionally avoided directing or highlighting head movement. By coupling this strategy with that of removing the study from a laboratory environment it was hoped that participants would be more likely to display head movements only when they felt internally motivated to do so.

In addition to detecting differences of head movement behaviour between sighted and visually impaired groups, the measurements of head movement in the field study could also provide baseline information for implementing head movement in virtual reality. For instance, if a system were to be designed using manual control input rather than head tracking to initiate simulated head movements, then information regarding the frequency/rates of movement in yaw, roll and pitch may be useful when determining desirable control sensitivities[42].

6.1.2 Hypotheses

The hypotheses under test were:

- In the absence of visual perception, head movements outside of postural oscillations (described in *chapter 5.3.1*) are measurable in visually impaired participants during navigation

- The quantitative and qualitative use of head movement will vary significantly between sighted (S) and visually impaired (VI) participants at critical points during test route navigation (described in *chapter 6.2.1*) ($H_1\colon \mu_{VI} \neq \mu_S$)

---

[42] The property of a controller that determines an appropriate output in response to a given input, in this case an appropriate rate of simulated head movement in response to the user’s manipulation of a device such as a joystick

6.2 Methodology

The head movement field study was conducted using a specially developed sensor toolkit described in *chapter 5,* which was used to record head movement, sound, and image whilst participants navigated a pre-defined route through an otherwise uncontrolled environment.

6.2.1 Route Selection

A short route through an outdoor urban environment was required for the study. When selecting the route, it was important to consider both the behaviour that navigating this route was likely to elicit from test participants, as well as the performance of the sensor package.

An urban environment was chosen at it would expose the test participants to complex combinations of external factors as they navigated the route:

- Sound emitting entities such as vehicles and other pedestrians
- Sound reflecting entities such as architecture

Although predefined, the route would pose the same conditions as any visually impaired individual would face when navigating in an outdoor, built environment to a fixed destination.

To simplify analysis of the head tracker data, a "one way" route was used, such that the participant would follow the same approximate bearing throughout the entire route. This would simplify head tracking data by preventing reversals of the tracker coordinate which represented 0º of yaw displacement.

The route selected started at the Queens Building, in De Montfort University's campus, and proceeded through a pedestrianised area, to a public main road. The route included three road crossings, two of which were open to vehicular traffic, and was bounded by ten large adjacent buildings at varying distances from the route. Participants were instructed to use pedestrian crossings freely, for safety reasons. The route was a total of ≈272m long.

Importantly, each road which participants were required to cross included tactile pavements at the kerb. In general, such pavement surfaces are used to indicate the presence of a relatively safe crossing position to visually impaired pedestrians. In the context of the study it also meant that visually impaired participants would be aware of roads, even those closed to vehicular traffic. This feature would prove useful for test analysis, where road crossing behaviour in the presence of traffic sound could be compared to that without traffic sound.

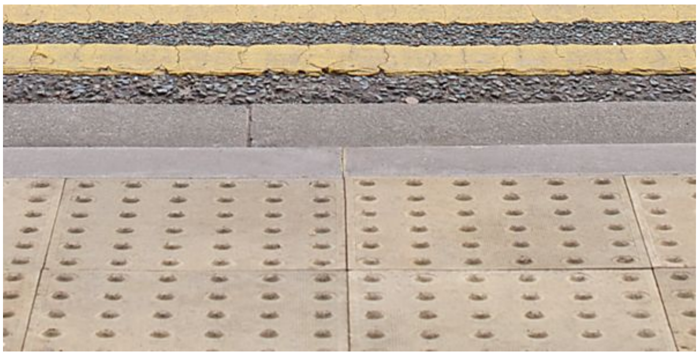


Figure 6.a An example of tactile paving on the test route, placed to signal the presence of roads and crossings to VI pedestrians

6.2.2 Equipment Preparation

Prior to the arrival of each participant, a new folder was created to separately store tracking data, video, and audio recorded by the test toolkit.

After the participant briefing, a hat containing the inertial tracker, microphones, and webcam was placed upon the participant's head (this portion of the sensor kit weighed approximately 150g). The participant was then turned to face in the direction required to begin navigating the test route and instructed to stand in a relaxed position with their head straight, facing forwards. The toolkit was then initialised meaning that:

- The inertial sensors were zeroed to the participant's current head orientation
- Audio, video, and sensor data began to record
- Initialisation signals (as described in *5.2.2)* were triggered for audio, video, and sensor data synchronisation

Participants were then asked to complete the task which had been described to them during the pre-study briefing.

6.2.3 Participant Briefing

To achieve a blind study, participants were not informed of the exact nature of the experiment, or its precise topic of interest, until after they had navigated the test route. Participants were however, aware that the equipment they were carrying could sense motion, as well as record sound and images.

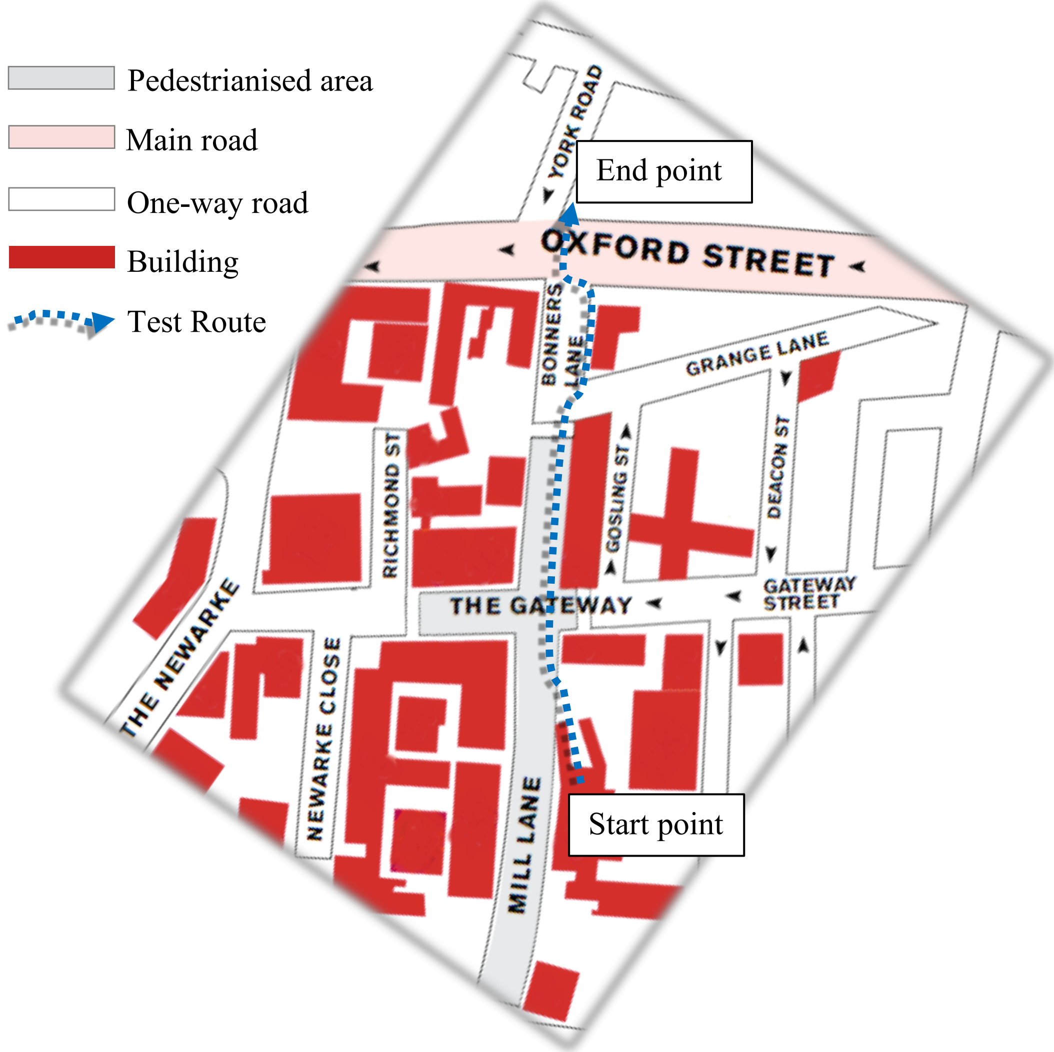


Figure 6.b Map of the test route selected for the field study, with start and end points marked

Instructions were given verbally to participants, from a pre-written script (available in *Appendix Two)*, to ensure that all participants received the same instructions regardless of their sighted or visually impaired status. The map of the test route in *figure 6.b* was shown to them, along with a verbal description of the route. It was emphasized that participants should navigate the route as if it were any other journey, using whatever safety precautions they would normally use whilst walking in similar environments.

6.2.4 Post-Study Debriefing

Participants were met by the researcher at the end of the navigation route, whereupon the sensor package was removed and data recording was stopped. Participants were then led back to the start point of the route for a post-test debriefing.

At this stage participants were informed of the exact nature and topic of the study and were invited to share information regarding their habits when navigating in urban environments, both during the test, and in general. Although not a formal questionnaire, it was hoped that participants may offer information of interest to the research; elucidating their own test results or offering new avenues for investigation. The qualitative responses gathered during debriefings are discussed in context throughout the evaluation of results (from *6.3* onwards).

6.2.5 Study Sample Groups

Due to the nature of the study, several ethical concerns were addressed prior to selecting the study group. In addition to the usual concerns for data protection, informed consent, and so on, this study presented the following ethical concerns (the suitability to which participants were screened[43] prior to the commencement of testing):

- The toolkit that participants were required to carry throughout the study weighed ≈3kg in total (including a head mounted portion, weighing ≈190g)
- To attempt to isolate the participant's behaviours to only those related to navigation, the study required participants to participate without the direct influence/assistance of others
- The study required participants to walk a route of ≈272m which included crossing a road open to traffic

[43] Via information supplied with the call for volunteers, as well as a pre-study questionnaire and informed consent

- Since the study was primarily concerned with head movement, participants with neck/spine injuries or pathologies limiting movement were deemed unsuitable for study;
- Individuals with hearing impairment were similarly considered unsuitable for study

Participants were recruited via notifications placed in VISTA[44] news bulletins and De Montfort University internal email bulletins. A total of eight participants, four sighted, and four visually impaired participated in the study. The visually impaired group included two male and two female participants of ages ranging from 22 to 69. The sighted group consisted of two male and two female participants of ages ranging from 21 to 64. Of the visually impaired group three participants had early-onset of visual impairment, and one had late onset (ongoing for more than 10 years).

6.2.6 Statistical Methods

Owing to the small study group, it was inappropriate to divide the visually impaired group between early and late onset sight loss for quantitative analysis, so the groups were treated as a whole. The inertial sensor data from each participant was first analysed for average head movement (yaw, roll and pitch) frequency. Since head movements were represented by oscillations in sensor data over time, the values representing yaw, pitch, and roll displacement were isolated and normalised such that all magnitude values representing the extent of head displacement were constrained between 0 and 1 for the purposes of frequency detection.

The sensor data was filtered using a 6th order Butterworth filter, with a corner frequency of 3Hz for yaw, roll, and pitch, to remove data pertaining to rapid, passive head movements arising from locomotion, whilst retaining the slower deliberate head movements and those related to eye-head coordination in attention focusing [231] [233] [234]. The remaining peaks in data, representing yaw, pitch and roll displacements, were then analysed to determine the total number of head movements performed by each

---

[44] A charity supporting visually impaired individuals in Leicestershire and Rutland, UK

participant during the test. The velocity of turns was also retrievable from this data. Power spectral density calculations were made in order to establish the ranges dominant head turn frequencies used by participants during the test.

Unpaired t-tests with confidence interval (CI) of 95% were used to assess the significance of differences between the sighted and visually impaired groups. The standard deviation of results from the groups tended to be similar, indicating that a t-test would be viable, although limited in power.

Some qualitative assessments were made in cases where the quantitative analysis software (described in chapter *5*) was unable to identify certain trends in data. This was most commonly used when analysing data from isolated areas within the test route, for example when assessing behaviour at a particular road crossing, where head movement data in some planes was insufficient in frequency to appear in a Fourier transform of the data. In these cases, a simple plot of head angle over time was created to establish whether angular displacement had occurred, and video/audio recordings from the appropriate section of the test route were manually reviewed to establish plausible reasons for such movements to exist (tracking traffic movement vs. general head sweeping), and to compare them between participants and groups, again using unpaired t-tests with confidence interval (CI) of 95%.

6.3 Discussion of Average Head Turn Regularity

The initial quantitative analyses of head turn behaviour across the duration for the entire test route, as well as of the time each participant took to complete the route yielded no significant variations between groups, however some qualitative differences between the performance of participants was noted.

Comparison of sighted and visually impaired participant's total head movement during study							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
VI 1	7:33	339	114	278	.74	.25	.63
VI 2	3:58	124	19	3	.52	.08	.01
VI 3	9:06	335	33	22	.61	.06	.04
VI 4	5:44	193	0	0	.56	0	0
Visually Impaired Group Mean					.61	.09	.17
Standard Deviation					.09	.09	.30
S 1	5:26	189	0	7	.57	0	.02
S 2	6:55	236	33	19	.56	.08	.04
S 3	6:09	221	68	15	.59	.18	.04
S 4	6:14	232	87	9	.62	.23	.02
Sighted Group Mean					.58	.12	.03
Standard Deviation					.02	.10	.009

Figure 6.c Average head movement frequency for sighted and VI participants during the test

Differences in the average number of heads turns per second between visually impaired and sighted groups over the entirety of the test route were non-significant: (CI=95%; yaw: t(6)=.45, P=.66; pitch: t(6)=.33, p=.74; roll: t(6)=.91, p=.39).

In general, the largest performance differences between visually impaired and sighted groups arose not from the frequency of head turns used during navigation, but from the time individual participants in each group required to complete the test route. Although the differences in completion time for each group were not statistically significant (mean completion time:
VI = 6:30, S = 6:11; SD: VI = 2.3, S = 0.5; CI = 95%; t(6) = 0.28, p = .78), the range of results in visually impaired participants was much larger than that of sighted participants, reflected by the difference in standard deviation of completion times. This variance may be partially accounted for by the participant's familiarity with the test route; participant

VI2 (test duration = 3:58) is a long-term employee of De Montfort University, and regularly navigates its campus (including the area in which the test route was based). During post-test discussion, the participant revealed that they were familiar enough with the area to determine their approximate location by counting steps as they walked- possibly leading to a reduced need for auditory evaluation of their environment. In comparing the completion times of VI2 with those of the other visually impaired participants, it became apparent that both VI2 and VI3 (a participant unfamiliar with the test route prior to testing) had encountered the lowest waiting times at road crossings. In total VI3 spent ≈30s longer waiting at crossings than VI2, however their total test time was ≈5.08 longer. Although far from conclusive, this may give some indication as to the importance of route/environmental familiarity for ease of navigation in the absence of vision, and to one viable strategy for cognitive mapping in the visually impaired.

Although two visually impaired participants demonstrated a high total number of head movements during the test (VI 1 = 339 and VI 3 = 335), they required larger amounts of time to complete the test route when compared to both visually impaired and sighted participants studied. Further examination of participant VI 3's behaviour during the study revealed that they had paused at the tactile pavement of the closed road and used yawing head movements for longer before crossing, when compared to other visually impaired participants. They also took the longest of any participant to navigate the pedestrian sections of the test route. Given that this participant was the eldest of all participants, it is not unexpected that they walked more slowly through the route, however their apparent caution at road crossings is indicative of variations in behaviour which could hypothetically have arisen in any participant- although they reported confidence in navigation, and road crossings, during the consent and briefing stages of the study, they may have developed a variety of personal strategies (not related to head movement) in order to ensure their own safety in such scenarios.

6.4 Discussion of Head Movement Velocity

The inertial sensor data was analysed for head turn velocity. This analysis aimed to assess the average rates of head rotation throughout the entire test route, rather than the regularity with which it occurred. As in the previous analysis, the magnitude data was first normalised, a periodogram analysis of mean head turn frequency was then conducted to determine the most significant frequencies (and therefore velocities in radians per second) recorded for each group of test participants.

Mean yaw turning rates in the sighted group were concentrated at ≈1.5 radians, or 85º per second. In the visually impaired group, a far greater range of significant head turn velocities was recorded, most notably concentrated between ≈3.14 radians, or 179º per second and up to ≈9.42 radians, or 539 º per second:
(VI = 5.4rad/s, S = 1.5rad/s; SD: VI = 2.9, S = 0.54; CI = 95%; $t(6) = 2.5$, $p = .04$).

Despite the non-significant differences in total instances of head movements, the visually impaired group displayed more rapid yaw head movements than their sighted counterparts. This was largely associated with the fact that they tended to perform wider, sweeping motions of the head, as oppose to the sighted group whose head movements tended to be guided towards specific environmental features such as traffic lights and oncoming traffic. Pitch and roll head movements were relatively slower for all participants, with typical movement velocities falling between ≈.12 radians, or 7º and ≈1.57 radians, 90º per second.

The only significant quantitative differences between sighted and visually impaired participants existed in head movement velocity for yawing movements, there were notable qualitative differences in the use of head movement between groups. These were particularly pronounced at sections of the test route which involved road crossings, and are discussed in line with the information presented about these sections throughout *chapter 6.5.* Nevertheless, the visually impaired group did display head movements beyond those associated with postural oscillation, supporting the hypothesis that they would.

6.5 Comparison of Head Movement in Road Crossing Scenarios

A final analysis for head turn behaviour was conducting, comparing crossings with or without the presence of traffic for both groups, and a section of pedestrian only pavement.

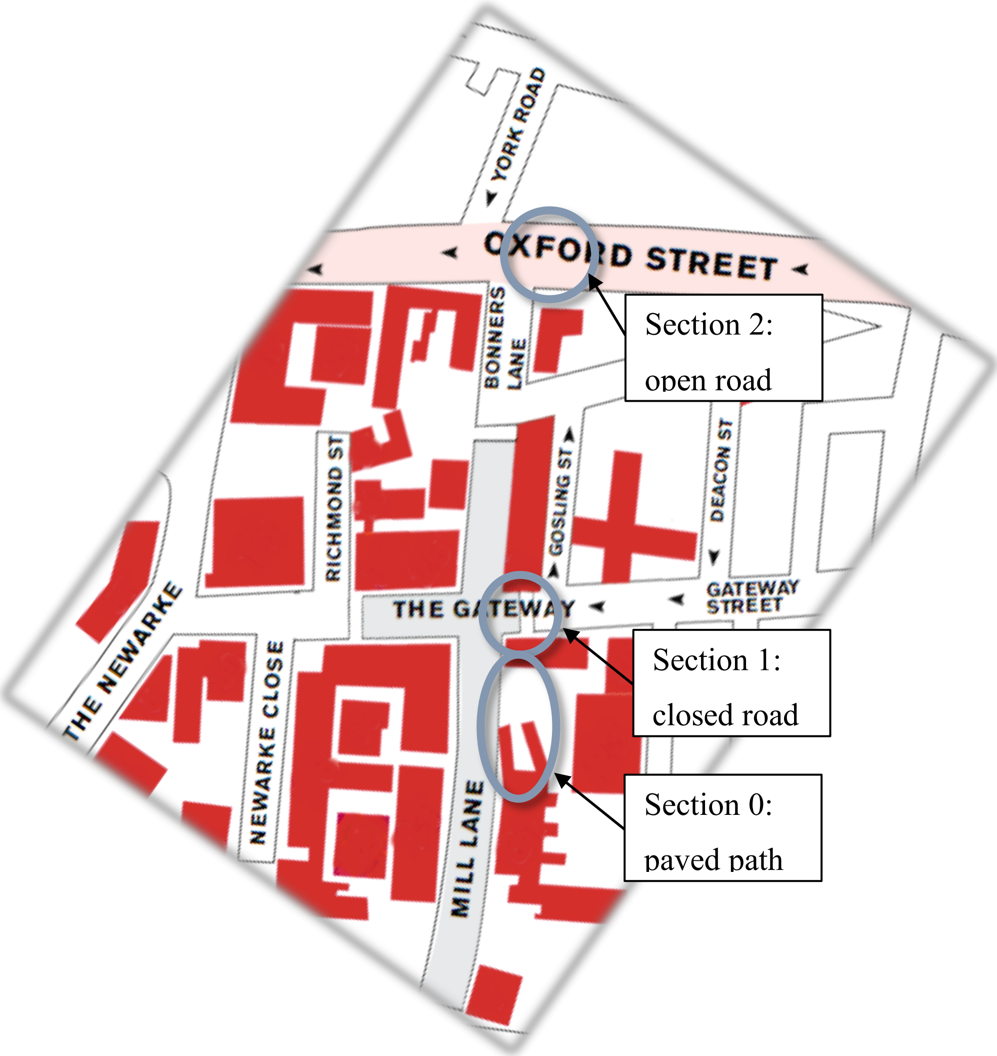


Figure 6.d A map showing the locations of the open and closed roads, used in the final head movement data analysis

In this analysis the normalised, low pass filtered, magnitude data was isolated for 2 specific sections of the test route, which incorporated a closed, and open road crossing.

The analysis comparted the behaviours of sighted and visually impaired participants at each of these crossings and compared differences within each group. Although the analysis software described in *chapter 5* was used for the initial quantitative analyses, some manual verification of tracking data and images captured during the test was employed to conduct a qualitative analysis of behaviours. This became especially useful when examining pitch and roll behaviours of the participants, where interesting differences in the apparent motivations for pitching or rolling the head were noted.

6.5.1 *Comparison of VI and Sighted Participants at each Crossing*

The total number of head turns, and average head turn frequency was calculated at both crossings for each group. A mean head turn frequency for each group was also determined and analysed for statistical significance at each of the route sections.

The first analysis was that of sighted and visually impaired participants at section 1: closed road.

Comparison of sighted and visually impaired participant's head movement at section 1: closed road crossing							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
VI 1	9.6s	8	0	0	.74	0	0
VI 2	10.9s	8	0	0	.52	0	0
VI 3	24.8s	15	0	0	.61	0	0
VI 4	13.7s	6	0	0	.56	0	0
Visually Impaired Group Mean					.61	0	0
Standard Deviation					.09	0	0
S 1	9.9s	7	0	0	.57	0	0
S 2	11.9s	8	0	0	.56	0	0
S 3	13.8s	9	0	0	.59	0	0
S 4	13.7s	6	0	0	.62	0	0
Sighted Group Mean					.58	0	0
Standard Deviation					.02	0	0

Figure 6.e Comparison of participant head movements at section 1 of the test route, with group mean and standard deviations shown

In section 1, the closed road portion of the test route, there was a difference in variance between sighted and visually impaired groups (yaw: $t(6) = .45$, $p = .66$) with sighted participants showing lower mean head turns per second than their visually impaired counterparts, but with a far smaller range in mean head movements. No member of either group appeared to display pitch or roll head movements.

Next, the behaviour of the two groups at section 2: open road, was analysed.

Comparison of sighted and visually impaired participant's head movement at section 2: open road crossing							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
VI 1	104.2s	84	4	0	.8	.03	.03
VI 2	16s	8	0	0	.5	0	0
VI 3	138.5s	91	2	0	.65	.01	.02
VI 4	39.3s	27	0	0	.68	0	0
Visually Impaired Group Mean					.65	.01	.01
Standard Deviation					.12	.01	.01
S 1	75.4s	35	0	0	.46	0	0
S 2	15.9s	8	0	0	.5	0	0
S 3	103.9s	54	3	0	.51	.02	0
S 4	43.4s	29	1	0	.66	.02	0
Sighted Group Mean					.53	.01	0
Standard Deviation					.08	.01	0

Figure 6.f Comparison of participant head movements at section 2 of the test route, with group mean and standard deviations shown

Section two showed no significant differences in mean head movement between groups (yaw: t(6) = 1.65, P = .14; pitch: t(6) = 0, p = 1; roll: t(6) = 1.6, p = .14). Upon manual inspection of the test data some qualitative differences were found for head movements at the open road crossing. The sighted group tended to use yawing movements to place either the crossing lights, or the road to their right (the direction of oncoming traffic) closer to 0º azimuth in the horizontal plane. Indicating that their attention was shifting between oncoming traffic and the crossing signal. The visually impaired participants showed less tendency to actively track traffic with head movement, instead performing more general sweeping patterns per second- independent of the movement of traffic (VI = 0.19, S = 0.04; SD: VI = 0.26, S = 0.01; CI = 95%; $t(6) = 10.2$, $p = <0.01$).

Next, the behaviour of the sighted group at each of the route sections was analysed.

Comparison of sighted participant head movement at section 1 & 2: closed vs. open road crossing							
Section 1: closed road crossing							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
S 1	9.9s	7	0	0	.7	0	0
S 2	11.9s	8	0	0	.67	0	0
S 3	13.8s	9	0	0	.65	0	0
S 4	13.7s	6	0	0	.43	0	0
Mean					.61	0	0
Standard Deviation					.12	0	0
Section 2: open road crossing							
S 1	75.4s	35	0	0	.46	0	0
S 2	15.9s	8	0	0	.5	0	0
S 3	103.9s	54	3	0	.51	.02	0
S 4	43.4s	29	1	0	.66	.02	0
Mean					.53	.01	0
Standard Deviation					.08	.01	0

Figure 6.g Comparison of only sighted participant's head movements at sections 1 & 2 of the test route, with group mean and standard deviations shown

No significant difference was found when comparing the head movement behaviour of sighted participants at the closed road compared to the open road

(yaw: $t(6) = 1$, $p = 0.33$; pitch: $t(6) = 1.7$, $p = .13$). although mean head movement frequency was slightly lower at the open road, in the presence of traffic.
The slight differences in pitch and roll movements at the closed and open roads pointed to an interesting, if somewhat predictable behaviour in the sighted participants; although the video captured during the test revealed that their attention appeared to be largely concentrated upon the road, and the traffic upon it, head movements directed towards the crossing lights were observable. In participants S3 and S4 these movements appeared to be slight, occasional, diversions from the more common yawing movements displayed whilst observing the road. These movements occurred less frequently in participants S1 and S2, who generally showed marginally fewer rapid head movement at the open road.

Lastly, a comparison of visually impaired participants at section one and two was conducted.

Comparison of visually impaired participant's head movement at section 1 & 2: closed vs. open road crossing							
Section 1: closed road crossing							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
VI 1	9.6s	8	0	0	.83	0	0
VI 2	10.9s	8	0	0	.73	0	0
VI 3	24.8s	15	0	0	.61	0	0
VI 4	13.7s	6	0	0	.46	0	0
Mean					.65	0	0
Standard Deviation					.15	0	0
Section 2: open road crossing							
VI 1	104.2s	84	4	4	.8	.03	.03
VI 2	16s	8	0	0	.5	0	0
VI 3	138.5s	91	2	3	.65	.01	.02
VI 4	39.3s	27	0	0	.68	0	0
Mean					.65	.01	.01
Standard Deviation					.12	.01	.01

Figure 6.h Comparison of only visually impaired participant's head movements at section 1 & 2 of the test route, with group mean and standard deviations shown

Like the sighted group, the visually impaired group showed no statistical difference in head movement frequency whether at a road with traffic, or without
(yaw: $t(6) = 0$, $P = 1$; pitch: $t(6) = 1.4$, $p = .2$; roll: $t(6) = 1.4$, $p = .2$).

In these tests, as in previous tests of head tracking data, very little pitch or roll oscillation was measurable through simple frequency analysis. A rather interesting contrast between pitch and roll movements was revealed upon manual examination of the test data, as well as a difference in general behaviour at the open road crossing. The visually impaired test participants appeared to prefer standing towards the outer edges of the textured pavement, closer to the location of the traffic light control boxes. Although possibly a mere coincidence, it appears sensible to suggest that this behaviour would be beneficial in terms of allowing visually impaired persons to easily hear audio signals produced by the traffic lights, indicating that it was safe to cross. It is also the case that many pedestrian crossing controls include a haptic feedback device for use by people with sensory impairments, although such a device was not present on the traffic lights in the test route, visually impaired persons may well develop a habit of standing closer to the control boxes so that they can use such a device if it happens to be present. In terms of head movement behaviour, much like their sighted counterparts, the visually impaired group did show slightly higher rates of pitch and roll at the open road crossing.

The only analyses which showed statistically significant differences were that which compared behaviour at the closed road crossing: the sighted group showed lower mean frequency head movement at the closed road, but with far less range between individual participants, and that which compared the qualitative behaviour of the two groups, where the visually impaired group showed more general head-sweeping motions when compared to the sighted group, who's head movement tended to be associated with tracking traffic.

Although it is impossible to determine exactly why the difference in mean head movement frequency arose, there are factors which may have influenced the behaviour of sighted test participants differently than their visually impaired counterparts, and which have been partially assessed from video footage recorded during the test:

- Participants at the closed road section do not experience any traffic noise, which may have encouraged some visually impaired participants to cross with less assessment of the road's safety
- There are visible bollards near the closed road, indicating that it is closed, approximately 10m to the right of the crossing point. This visible cue may have affected the sighted participants' behaviour in some way as they tended to display head movements between the direction of the bollards and the road ahead.

It is impossible to determine the exact focus of visual attention from participants as the tracking package did not include gaze detection, although it may seem reasonable to assume that visual stimulus would perhaps have greater influence on the behaviour of sighted participants compared to visually impaired participants, particularly given that the head movement of sighted participants at the open road was guided towards traffic.

It is also possible that head movement is used in a methodical way to evaluate the local sound scene by visually impaired individuals; in post-test debriefing both participants VI1 and VI3 indicated that they felt head movement was helpful to some extent when judging sound distance and direction for movement (both participants were affected by early onset sight-loss) but in the absence of sounds of interest, this behaviour may also be absent.

6.5.3 Comparison of Sighted and VI Participants in a Pedestrian Area

The final set of analyses of head tracking data were conducted for a pedestrian area of the route, containing no road/traffic crossing sections (indicated as section 0 in *figure 6.d*). This area is of interest as it contrasts from sections one and two in that participants would not expect to encounter fast moving hazards and, as it is situated within the larger pedestrianised area of the route, nearby auditory stimulus would exclude motor vehicles.

Comparison of sighted and visually impaired participant's head movement at section 0: pedestrian path with no crossing							
Participant	Test Duration	Total Head Movements			Average Head Movements/Second		
		Yaw	Pitch	Roll	Yaw	Pitch	Roll
VI 1	69s	41	1	0	.59	.01	0
VI 2	63s	33	0	1	.52	0	.01
VI 3	75s	46	0	0	.61	0	0
VI 4	65s	44	1	1	.67	.01	.01
Visually Impaired Group Mean					.59	.005	.005
Standard Deviation					.06	.005	.005
S 1	56s	47	0	1	.83	0	.01
S 2	67s	34	2	1	.5	.02	.01
S 3	58s	41	1	1	.7	.01	.01
S 4	68s	37	2	0	.54	.02	0
Sighted Group Mean					.64	.01	.007
Standard Deviation					.15	.009	.005

Figure 6.i Comparison of only visually impaired participant's head movements at section 0 of the test route, with group mean and standard deviations shown

Analysis of group means shows there are no significant differences between sighted and visually impaired participants (yaw: $t(6) = .54$, $P = .6$; pitch: $t(6) = 1.34$, $p = .2$; roll: $t(6) = .65$, $p = .5$) although there is a much wider range in mean yaw turns per second for the sighted group compared to the visually impaired group. It is interesting to note that there is likewise no significant difference in the times taken; without factors such as traffic to slow progress, both groups completed this ≈90m section of the route in similar time frames.

6.6 General Discussion

The hypotheses under test were:

- In the absence of visual perception, head movements outside of postural oscillations (described in *chapter 5.3.1*) are measurable in visually impaired participants during navigation

- The measured quantitative and observed qualitative use of head movement will vary significantly between sighted (S) and visually impaired (VI) participants at critical points during test route navigation (described further in *chapter 6.2.1*)

$(H_1: \mu_{VI} \neq \mu_S)$

No significant difference in mean head movement between visually impaired and sighted individuals was discovered during this experiment, perhaps because of the small sample size and deviations in individual behaviour within either sample group (the visually impaired sample tended to show larger ranges and deviations in mean head movements than the sighted group), or perhaps because no large difference exists in the population. Yawing velocities were different between sighted and visually impaired participants, such that the visually impaired demonstrated higher head turning velocities. Both sighted and visually impaired individuals actively use head movement during navigation- despite it being impossible to correlate this with auditory perception in the context of the results found here; it is interesting to note that two participants affected by early onset sight-loss consider head movement to be linked to auditory perception of movement. A number of studies have suggested differences between individuals with early and late onset sight-loss, in terms of neurological activity in auditory tasks, which may account for some differences in auditory perception between the two groups [235] [236] [237].

Qualitative differences in head movement were revealed, at the open road crossing the sighted group guided their attention towards traffic and the crossing signal, whilst the visually impaired group tended to perform more general sweeping patterns of head movement. It was also the case that the visually impaired chose to stand closer to the crossing signal control box whilst waiting to cross. In the absence of visual verification, this strategy would allow them to ensure that the crossing button had been pushed. It would also allow them to make use of accessible signals such as haptic and audio feedback from the crossing signal.

To move beyond these results and investigate the self-perceived use *and* possible role of head movement in auditory perception, a new study was devised in the form of a self-report questionnaire of the visually impaired (discussed in *chapter 7*).

Part III: Self-Report Questionnaire

7. Self-Report Questionnaire

This chapter details the design and analysis of a self-report questionnaire study of the visually impaired (both early[45] and late blind[46]) and sighted, including the aims of the study, and the sample selection (*chapter 6.1.1 & 6.1.2*), the questionnaire development (*chapter 6.2*) and the questionnaire results (*chapter 6.3*). The questionnaire aimed to gather qualitative data regarding navigation strategies employed by visually impaired individuals, with a view to understanding the prevalence and use of head movement to aid auditory scene evaluation (the full questionnaire is available in *Appendix Three*).

7.1 Rationale and Hypotheses

7.1.2 Rationale

The design of this questionnaire was guided by the prediction that participants may offer information that would support the understanding of whether the visually impaired reported using head movement in a way which they believed was significant to audition and navigation. By constructing a questionnaire, it was hoped that a larger sample could be tested: the greater physical safety of participants in a questionnaire compared to the previous observational study would raise less ethical need to exclude potential participants, and perhaps more people would be inclined to participate in a study which could be conducted via the telephone, excluding any need to travel to an unfamiliar location. Both quantitative (in the form of Likert scale responses) and qualitative (in the form of open-ended responses) data were gathered from participants, so that both data sets could be compared within and between groups, to indicate participant's reported confidence in navigation strategies versus their reported use of individual strategies.

By opening the study to a potentially wider group of participants it was possible to sub-divide the visually impaired group into individuals with early and late-onset sight-loss,

[45] In this case considered to be anybody registered blind, with onset at or before the age of five

[46] In this case considered to be anybody registered blind, with onset at or after the age of six

although the bootstrap statistical method was still used to generate larger volumes quantitative, re-sampled results.

Although the guiding focus of the study was to examine the role of head movement and angle specifically in auditory distance perception (for hypotheses see *chapter 7.1.3*), some exploratory questions regarding angular localisation, and sounds which presented particular localisation difficulty or ease were also included.

7.1.3 Hypotheses

The aims of the study were informed by data gathered from the field study discussed in *chapter 6*; although head movement was observed in visually impaired individuals during the navigation, the small sample group made it problematic to conclude that such results would be seen in the larger population, or indeed to firmly determine the use of head movement being related to auditory perception. As stated previously, two participants indicated that head movements may be associated with resolving ambiguities in sound source distance, particularly for moving/accelerating sound sources such as vehicles.

In developing the hypotheses of this study, it was taken that:

- The acoustical properties and activity of certain sound sources would make them more difficult to locate aurally than others [31] [30] [29].
- In order to safely navigate an environment, certain sound sources would be of greater interest than others.
- In the presence of complex sound fields with multiple sound sources, recognition of a sound source may be more difficult without the use of head movement [63] [64].
- Visually impaired individuals may posess some amount of functional eyesight [238] which may be used in conjunction with auditory perception to facilitate safe navigation of a given environment [213].
- Hypotheses would be tested using the factors identified in *chapter 7.2.4*
- The five-point Likert scales used to gather quantitative responses used lower values to indicate more positive responses.

The hypotheses were then:

- Visually impaired individuals (VI) would self-report greater auditory localisation confidence than sighted individuals (S) (when responding to factor 1):

$(H_1: \mu_{VI} < \mu_S)$

- Visually impaired individuals (VI) would self-report head angle and movement as a factor in auditory distance perception during navigation (number of VI responses to factor 1 at < 3 is greater than 40%[47] of VI total responses to factor 2)

- The self-reported use of head angle and movement to facilitate auditory perception in individuals with early onset sight-loss (E) will vary significantly from that of those with late-onset sight-loss (L), and sighted individuals (S) (when responding to factor 2):

$(H_1: \mu_E \neq \mu_L \ AND \ \mu_S)$

7.2 Methodology

Although self-report methods are known to be inferior to direct behavioural measurements in several ways (detailed further in chapter 7.2.2) the use of a questionnaire-based method was selected to eliminate certain ethical issues which had previously precluded some individuals from the study in chapter 6. Owing to the use of Likert scales for data gathering, quantitative responses in this study were limited to a finite and definite range, the bootstrap method (described in chapter 7.2.6) was viable as a means to validate the normality of data and generate a representative re-sampled set of data.

The remainder of the self-report questionnaire was concerned with qualitative data gathering, providing participants with open-ended questions in which they could

[47] Chance level

elaborate upon quantitative responses given, and report other information which they felt pertinent. Notes on the general structure/design of the questionnaire and how it avoided leading participants can be found in chapter 7.2.3.

7.2.1 Self-Report Questionnaire Structure

The population from which the questionnaire sample was drawn was adults registered as visually impaired, and sighted adults. To make the questionnaire accessible to a large sample group whilst maintaining delivery control, it was administered by interview over the telephone. Participants were recruited through online, email, and print mail bulletins circulated via Action for Blind People[48], Vista News bulletin and De Montfort University electronic bulletin. Respondents willing to participate in the questionnaire were contacted directly by the interviewer[49] at an agreed time, the nature and purpose of the questionnaire was explained in accordance with guidelines provided by the De Montfort University research ethics committee. Demographic data including gender, age, and nature and duration of visual impairments was collected for each participant, but responses were anonymised by way of storing each individual's questionnaire responses under a participant number; the individual's names were not stored with any data collected from them.

Aside from demographic questions, and those questions relating to the respondent's use of auditory or visual perception as their predominant means of navigation, the questionnaire was divided into two sections. The first part concerned the participant's self-reported use of auditory and visual perception to understand their environment, and confidence in using auditory perception for navigational tasks. The second part concerned respondent's self-reported awareness of whether head angle tended to influence their ability to judge the location or movement of sound sources, and whether they used head movement to facilitate auditory perception. These sections were both presented as Likert

---

[48] A national charity working with the Royal National Institute for the Blind (now merged with RNIB) to support the visually impaired, and their friends, family and carers in the UK

[49] Which, in all cases, was the author of this document

scales with a maximum of five response options. Available response options ranged from strongly positive to strongly negative, as well as an uncertain/non-response option;

- Example of section one question: “Do you find it easy to tell in what direction a vehicle is by its sound?”

□Usually easy, □Sometimes easy, □Often difficult, □Usually difficult, □Unsure”

- Example of section two question: “Do you ever intentionally move your head to help you tell how far away something is by its sound?

□Always, □Sometimes, □Not often, □ Rarely, □Never ”

The Likert scale responses were translated into numerical responses, in which 1 indicated the most positive, and 5 the least positive or least sure response. This allowed for quantitative statistical methods (ANOVA) to be applied when analysing the responses [239], which in turn allowed for the use of bootstrapping (discussed further from *chapter 7.2.4 – 7.2.9*). Whilst there is some controversy of the appropriate number of points to use in Likert scales, the main statistical difference between 4-, 5-, 6- and 11-point scales is in the skewness of the responses [240]. Since ANOVA methods are generally robust to skewness [239] 5-point Likert scales with no mid-point, but instead an unsure/thoroughly negative option were employed (where the selection of this option was treated as thoroughly negative for the purposes of analysis), thus minimising the social desirability effect [241].

The purpose of separating the questionnaire into sections regarding respondents’ self-reported confidence in auditory perception and use of head movement was to assess if there were any correlations between increased confidence, and awareness of the use of head movement.

Open ended and categorical questions were included to gather qualitative data regarding respondent’s experiences of navigating in urban environments, including issues such as confounding/interfering sounds, and techniques used to determine the distance, movement, and position of sound sources within the environment. At the end of the questionnaire respondents that reported using head movement to facilitate audition in some way were asked to report on how they believed it helped them (if at all);

- Example of categorical question asked prior to any mention of head movement in the questionnaire process: "Do you normally use any physical actions or behaviours to help you tell if a sound is moving towards or away from you?"

- Example of open-ended question asked at the end of the questionnaire (after questions regarding self-reported use of head movement): "If you do find that moving your head helps you to tell how far away something is, or which direction it is moving by its sound, could you describe how you feel it helps?"

7.2.2 On Issues Arising from the Self-Report Method

Two key issues which generally weaken self-report methods compared to observational methods of behavioural study are, simply put: people cannot report accurately on that which they are unconscious of, and (more importantly) people are not generally conscious of higher cognitive processes, nor do they often realise that they are unconscious of them. Nisbett & Wilson roundly highlighted these issues [242]. Such issues do not automatically exclude self-report as a viable source of insight into behaviours and their motivational stimuli; in cases where the stimuli are both a salient and probable motivator for behaviour, and where a participant has established a personal knowledge of the stimuli and response from experience, it is possible for them to identify a causal relationship [242] particularly in self-focused study [243]. In the present case, this does not mean that individual participants could access the process by which head movement might assist in auditory localisation, nor does it mean that they would necessarily be correct in any assertions that it *does* or *does not* help, however it is possible that in a situation where their attention is outwardly directed to environmental stimuli they could identify patterns in their own responsive behaviour- particularly any voluntary behaviours. They may at least be aware of voluntary head movements which, although not necessarily initiated by a conscious process, would still have been "filtered" through one [244].

Considering the design of the questionnaire detailed in the remainder of this section, participants were given instructions to recall past experiences where auditory stimuli would be present. They were specifically directed to consider sound emissions when answering:

"When answering the following questions please try to think about times when you've had to walk or find your way in environments such streets or town/city centres… The following questions are about things that make sounds (for example vehicles, or people)"

The questions themselves were also directed towards the participant and sound sources and their nature:

"Do you find it easy to tell if a vehicle is moving towards or away from you by its sound?"

Such questions were proceeded with open questions about behaviour:

"Do you use any physical actions or behaviours to help you tell if a vehicle is moving towards or away from you by its sound (briefly describe them below)?"

Direct questions regarding head angle and movement were not presented until the final section of the questionnaire, at which point participants had already been offered the opportunity to freely answer questions about behaviour in general. Since the questionnaire was delivered by the researcher, participants were prevented from taking questions regarding head movement as a cue to revisit previous answers. Since the participants chosen were different from those included in the field-study, the possibility that results were contaminated by the participant's understanding of the expectations of the study was minimised. Although bias from past training which guided head movement was a possibility, no such training methods were reported by participants in response to:

"If you have learned any other skills to do with helping you understand the environment you are in by sound, please use the box below to tell us about them:"

To reiterate a note of caution regarding the results of the self-report study presented from *chapter 7.3* onwards: they do not amount to an objective measurement of the behaviours in question, but are a subjective assessment of them offered by the individuals who participated. Although it is fair to say that individuals may be aware of their own behaviour, and have an interpretation of what motivates it, it is not fair to say that such an interpretation is truly representative of any underlying neurological or psychoacoustic phenomenon by which auditory perception is achieved. Furthermore, the self-reports are still fallible to issues arising as a result of memory and previous instruction- it is possible for behaviour to be under-reported due to an absence of it in memory [245].

7.2.3 Study Sample Groups

Participants were recruited via news bulletins via Action for Blind People, and participants form the study in *chapter 6* were excluded form repeated study. A sample of 26 (16 male, and 10 female) participants between the ages of 21 and 72, with a mean age of 48 and a median age of 48.5, were questioned. The participants were organised into sighted (n = 13), early-onset (n = 9), and late-onset (n = 4) visually impaired groups. The sighted group were added retrospectively, following the surveying of the visually impaired groups. Of the visually impaired group (VI) 9 reported early onset of sight loss and four reported late onsets, with seven of those participants reporting total sight loss, and six reporting low vision.

Of these participants, one reported being affected by minor bilateral tinnitus, with all other participants reported no hearing impairments. This subject remained in the study owing to the nature of the symptoms self-reported[50]. No members of the sighted group reported any hearing related issues.

Since auditory perception was the primary concern of the study, individuals reporting profound hearing loss were excluded from the sample.

---

[50] A high frequency sound, self-reported by the participant as only audible in quiet environments such as “in bed, at night”

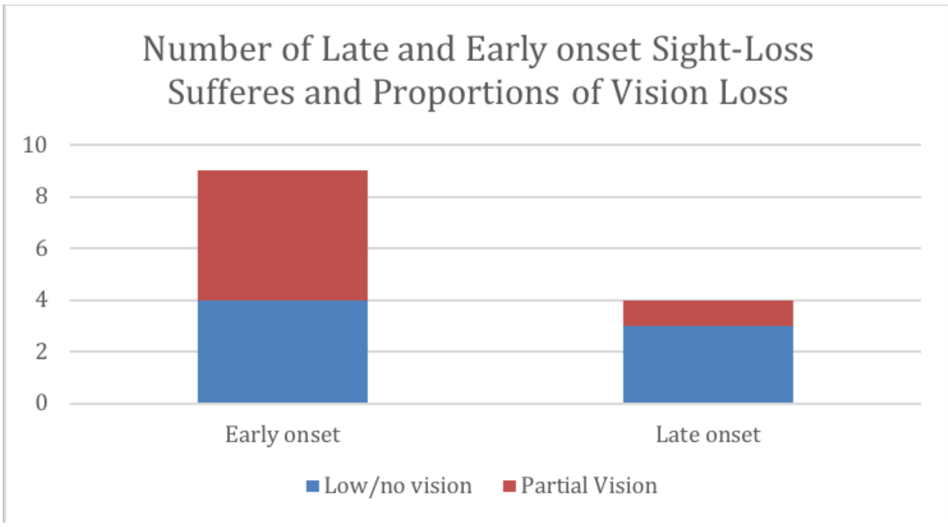


Figure 7.a Chart showing the relative proportions of study participants with early (left, n = 9), vs. late onset (right, n = 4) sight loss and severity of sight loss.

Ten visually impaired participants reported that they relied mainly upon hearing to understand occurrences in their local environment, with 3 reporting that they relied upon both hearing and vision. No visually impaired participants reported relying primarily upon vision. By contrast, five sighted participants reported relying mainly upon vision, and 5 reported relying upon both vision and hearing.

As the data collected was of an ordinal type (owing to the use of Likert scales, described in *chapter 7.2.1*) it was possible to model a larger population sample for analysis, using the Bootstrapping technique (described further from *7.2.6*). This meant that the modelled sample population consisted of 1,000 resamples with 27,000 complete, resampled, study responses in equally balanced groups.

7.2.4 Factor Analysis

To determine the factor structure, a principle component factor extraction was conducted upon the quantitative responses to the self-report questionnaire. Although the sample size was small, the data returned a poor, but not unacceptable [246] adequacy (KMO = .590) and a significant Bartlett's test of sphericity result (χ^2 (55) = 141.002, $p < .001$), indicating the data was suitable for factorability.

In total, 11 components (the quantitative questions form the questionnaire) were first examined and the commonalities were all above .5 (shown in *table 7.b*).

The initial eigenvalues showed that the first factor explained 34% of the variance, the second factor 26%, and a third factor 10%. The fourth, fifth and sixth factors had eigenvalues of just under one, explaining 8%, 6% and 3% of the variance respectively. A three-factor solution was explored, as 70% of the variance was explained by the first three factors, and each question showed a factor loading of at least .5 with one of these factors.

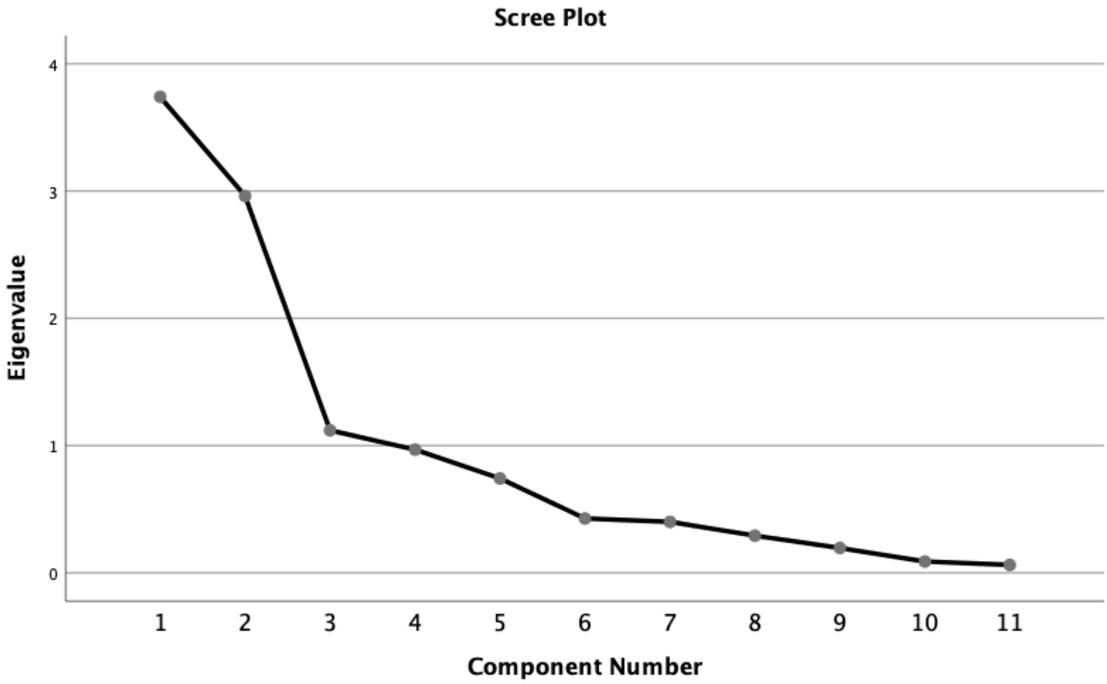


Figure 7.b Scree plot showing the eigenvalues for each component of the factor analysis. Components relate to questions 15, 16, 17, 23, 24, 25, 26, 31, 32, 33 & 34 (in ascending numerical order) in the questionnaire.

Five of the 11 questions (Q16, Q26, Q27, Q33, Q34- components 2, 6, 7, 10 and 11 in *figure 7.b*) showed loadings above .4 for two factors however, after varimax rotation with Kaiser Normalisation was applied each question correlated with only a single factor:

	Factor 1	Factor 2	Factor 3	Communalities
Q15		.683		.510
Q16			.832	.701
Q17		.837		.777
Q23		.788		.668
Q24			.717	.652
Q25		.875		.821
Q26			.693	.585
Q31	.856			.751
Q32	.877			.771
Q33	.917			.868
Q34	.824			.718

Table 7.b The factor loadings and commonalities of the 11 quantitative questions applied in the self-report questionnaire. Factor loadings shown are the result of principal factor analysis after varimax rotation with Kaiser Normalisation, with values < .3 supressed.

When comparing the questions according to their loading to each factor, the relationships between question themes were not entirely consistent:

Factor	Question
Factor 1	"Do you find that the angle of your head affects how well you are able to tell if something is moving towards or away from you by its sound?" "Do you ever intentionally move your head to help you tell if something is moving towards or away from you by its sound?" "Do you find that the angle of your head affects how well you are able to tell how far away something is by its sound?" "Do you ever intentionally move your head to help you tell how far away something is by its sound?"
Factor 2	"In general, do you find it easy to tell what direction something is in by the sound it makes?" "Do you find it easy to tell what direction a vehicle is in by the sound it makes?" "Do you find it easy to tell if a vehicle is moving towards or away from you by the sound it makes?" "In general, do you find it easy to tell if something is moving towards or away from you by the sound it makes?"
Factor 3	"Do you find it easy to tell how many vehicles there are by the sound they make?" "Do you find it easy to tell how far away a vehicle is by the sound it makes?" "In general, do you find it easy to tell how far away something is by the sound is makes?"

Table 7.c Questions grouped according to the initial 3 factors, with inconsistent question themes for factors 2 and 3.

Since the questions whose loadings correlated with the second and third factors broadly fit the theme of ease of localisation, a second factor analysis was conducted to explore the possibility of a two-factor solution. In this case, factor 1 accounted for 34% of the total variance, with factor 2 accounting for 26%.

	Factor 1	Factor 2	Communalities
Q15	.669		.452
Q16	.539		.301
Q17	.847		.718
Q23	.744	.	.584
Q24	.703		.509
Q25	.789		.684
Q26	.624		.404
Q31		.839	.703
Q32		.873	.766
Q33		.929	.863
Q34		.830	.718

Table 7.d The loadings and commonalities of the 11 quantitative questions applied in the self-report questionnaire for a two-factor solution. Factor loadings shown are the result of principal factor analysis after varimax rotation with Kaiser Normalisation, with values < .3 supressed.

In the two-factor solution, question 2:

“In general, do you find it easy to tell how far away something is by the sound is makes?”

Showed very low communality (.301). Since question 24 also dealt with auditory distance perception:

“Do you find it easy to tell how far away a vehicle is by the sound it makes?”

And its loading correlated with the same factor as question 2 and had a communality of .501, question 16 was removed from further analyses. The remaining 10 questions were then compared according to their correlations to each of the factors, and the factors were labelled:

Factor	Question
Factor 1: Self-reported auditory localisation confidence	"In general, do you find it easy to tell what direction something is in by the sound it makes?" "Do you find it easy to tell what direction a vehicle is in by the sound it makes?" "Do you find it easy to tell if a vehicle is moving towards or away from you by the sound it makes?" "In general, do you find it easy to tell if something is moving towards or away from you by the sound it makes?" "Do you find it easy to tell how many vehicles there are by the sound they make?" "Do you find it easy to tell how far away a vehicle is by the sound it makes?"
Factor 2: Self-reported use of head movement and impact of head angle	"Do you find that the angle of your head affects how well you are able to tell if something is moving towards or away from you by its sound?" "Do you ever intentionally move your head to help you tell if something is moving towards or away from you by its sound?" "Do you find that the angle of your head affects how well you are able to tell how far away something is by its sound?" "Do you ever intentionally move your head to help you tell how far away something is by its sound?"

Table 7.e Questions grouped according to the two-factor solution, with factor labels.

The two-factor solution yielded more consistent groupings of questions by theme, and so the factors proposed were retained for further analyses. Scores for each of the factors were computed; as the original responses were each given on five-point Likert scales, simple mean responses to questions which had their primary loadings on each factor were used.

7.2.5 Quantitative Statistical Analysis

Data from the composite scores of the quantitative questions was analysed using one-way ANOVA tests on Bootstrapped means, with a confidence interval of 95%. The bootstrapping technique is employed when inferences about a population need to be made based upon results from a small sample of that population. The technique was first proposed in 1979 by Bradley Efron [247] and is based upon a process known as resampling. Since its introduction bootstrapping has gained wide popularity, even having been referred to as a "gold standard" in popular statistics [248]. Soon after Efron's proposal, Singh further demonstrated the efficacy of bootstrapping [249] and Bickel & Freedman [250]. In this study, the quantitative statistical analyses are derived from Bootstrapped data.

7.2.6 The Principle of Bootstrapping

In statistical methods, the principle of plug-ins (where an unknown parameter can be substituted for with an estimate) is common. For example, when the population standard deviation is unknown, the standard deviation of a sample is used as an estimate in its place. In the case of bootstrapping however, all sample parameters are estimated [251]. The implications of this are that the bootstrapped parameters are not actually used as a substitute for the population parameters, but rather as an estimate for a larger sample parameter, based upon the parameters of the original sample. The bootstrapped statistics can be used to assess the accuracy of the original sample parameters, rather than to replace it.

7.2.7 Sampling and Resampling

Resampling forms the basis for implementation of bootstrapping and, as will become clear in this section, is where perhaps the largest assumption regarding this type of statistical analysis lies: when bootstrapping it is assumed that the data originally sampled from a population contains a similar range of responses as one would expect to find within the population as a whole. If sampling is simply the practice of selecting a number of representatives from a population for testing, then resampling is simply repeatedly drawing a random selection from those samples to form a new sample set.

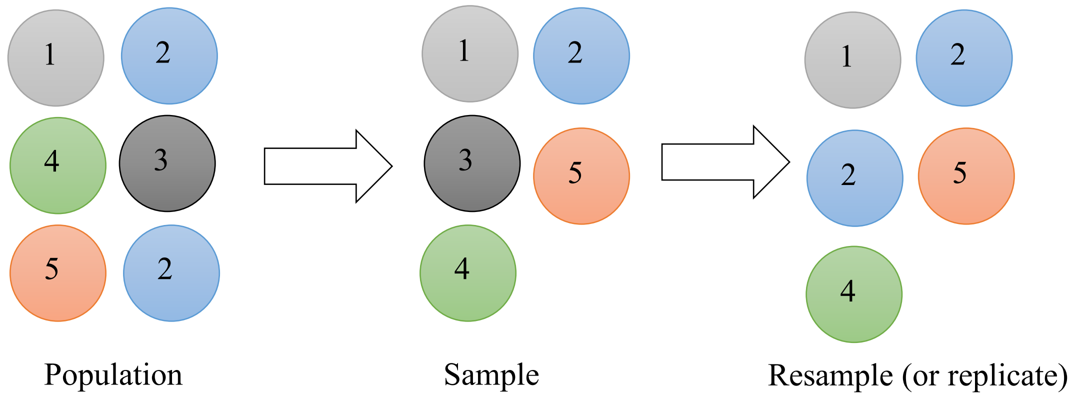


Figure 7.c Schematic representation of resampling in statistical analysis

It is important to note that the resample only ever draws from the original sample made, and never the actual population, and it is standard practice to draw many resamples, often reported as “replicates” of sampled data when used in conjunction with other statistical methods (such as ANOVA). The method of resampling is carried out with replacement, meaning that after a single response has been drawn into a resample selection, it may be drawn into it again. In this way, the resampled values may be a surrogate for larger sample populations; the cost of directly sampling a large population is eliminated whilst many estimated responses for that population can still be obtained.

7.2.8 Theoretical Support: Bootstrap with ANOVA

Much statistical analysis depends upon the Central Limit Theorem holding true; if a population parameter θ (such as a population mean response) is the subject of study, and a random sample of size n returns the data ($X_1, X_2, \ldots, X_n$) then the corresponding sample statistic of θ is $\hat{\theta}$. It is hoped (and indeed often assumed) that for a large sample size the distribution of $\hat{\theta}$ will be normal and centred on θ with standard error[51] $(\sigma/\sqrt{n})$, where a depends upon the population and the type of statistic $\hat{\theta}$. Although this may hold true for

[51] The standard deviation of sample means

larger sample sizes, in smaller samples it may be impossible to sufficiently demonstrate a normal distribution from the responses gathered. Bootstrapping offers a bypass for this issue by estimating a much larger sample size. Singh [249] and Bickel & Freedman [250] demonstrated the bootstrap sample distribution's asymptotic validity in most common statistics. The bootstrap distribution $\hat{\theta}_B$ is normal and centred at $\hat{\theta}$ with standard deviation $(\sigma/\sqrt{n})$, the distribution of $\hat{\theta}_B - \hat{\theta}$ approximates that of $\hat{\theta} - \theta$, forming the bootstrap central limit theorem.

To take advantage of the bootstrap method's power to estimate population parameters in conjunction with the more common ANOVA statistical method the quantitative inferential analyses presented here were treated by first performing a standard one-way ANOVA test to determine p and F values in the observed sample. The sample data was then resampled 1,000 times using the Bootstrap method (outlined in *7.2.6*) each resample, or replicate, was tested with the one-way ANOVA method to determine its p and F values. Mean p and F values were then calculated using the ANOVA results from the observed and Bootstrap replicated data.

This scheme, sometimes referred to as the ANOVA by simple bootstrap, is known to be more conservative (reducing the probability of Type I error) [252] and bootstrapping for ANOVA, as well as other forms of inferential statistical tests, is a generally accepted method [253] [254].

7.2.9 Self-Report Questionnaire Content Analysis

The qualitative data gathered via open-ended questions was organised by theme, based upon the question which was being responded to, and the actions and scenarios reported in participant's response to it.

The key thematic areas identified were:

- Descriptions of head movement patterns
- Discussion of sounds which elicited greater confidence in localisation
- Discussions of sounds which elicited reduced confidence in localisation

- Descriptions of scenarios in which sound localisation was problematic
- Descriptions of behaviours not related to head movement

Data was further organised into sub topics within each theme and repeated themes/topics are reported to indicate their popularity within the sample. The design of the questions meant that it was possible for a participant to reiterate a theme in two different questions, for example;

“Do you ever use any physical actions or behaviours to tell if a sound is moving towards or away from you?”

And

“Do you ever use any physical actions or behaviours to tell if a vehicle is moving towards or away from you?”

Could elicit the same response from a participant. In cases where it was clear that the participant *was* reiterating a previous answer, it was only counted once.

7.3 Self-Report Questionnaire: Quantitative Results

In this section, the results from the questionnaire are presented in terms of the two factors determined in *chapter 7.2.4*- Factor 1: Self-reported Auditory Localisation Confidence and Factor 2: Self-reported use of head movement and impact of head angle. Quantitative analyses are drawn from the Bootstrapped sample results, which are based upon the mean response values determined for each factor.

7.3.1 Factor 1: Self-reported Auditory Localisation Confidence

The bootstrapped mean response to factor 1 by group were:

- Early onset sight-loss: mean = 2.07, std. dev = .86, std. err = .28
- Late onset sight-loss: mean = 2.10, std. dev = .52, std. err = .23
- Sighted: mean = 1.60, std. dev = .58, std. err = .18

Although the mean responses for visually impaired groups were lower than those of the sighted group, the ANOVA showed no significant difference between groups:

Factor 1 ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.365	2	.682	1.418	.264
Within Groups	10.106	21	.481		
Total	11.471	23			

Table 7.f Results of one-way ANOVA for bootstrapped means of responses to Factor 1: Self-reported Auditory Localisation Confidence

The hypothesis that the visually impaired would report higher levels of auditory localisation confidence than the sighted is not supported by these results. For designers of auditory assistive technology, the result may still offer some encouragement. For the auditory localisation scenarios included in factor 1, there was a positive level of self-reported localisation confidence from the visually impaired groups.

Although it might have been expected that those affected by early onset sight loss would report higher confidence in auditory localisation (owing to some suggestions in literature that they possess improved localisation abilities when compared to late-onset, or sighted individuals), here it was in fact the sighted and late-onset blind individuals that appeared to report generally higher confidence. The lower reported confidence in visually impaired individuals when responding to the question regarding vehicles moving towards or away from them is very telling, particularly when considered in context with the qualitative responses provided (discussed in *chapter 7.4*). This perhaps reveals a more personal aspect to auditory localisation amongst the blind: the existence of some increased performance in auditory perception does not necessarily elicit higher levels of confidence in individuals which must rely upon it for their safety. Furthermore, even if an individual possesses such increased abilities, they will not automatically be aware of it.

7.3.2 Factor 2: Self-Reported use of Head Movement and Impact of Head Angle

In terms of the hypothesis that the visually impaired group would report the use of head movement and awareness of the impact of head angle above chance levels (40%), the responses to factor 2 indicated that the hypothesis could be accepted (66% of responses < 3, compared to the chance level of 40%).

In terms of group differences, the bootstrapped mean response to factor 1 by group were:

- Early onset sight-loss: mean = 1.97, std. dev = 1.01, std. err = .33
- Late onset sight-loss: mean = 2.75, std. dev = 1.63, std. err = .73
- Sighted: mean = 2.62, std. dev = 1.31, std. err = .41

Whilst the mean response for the early sight-loss group was lower than that of the other two groups, the ANOVA showed no significant differences:

Factor 2 ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.765	2	1.382	.839	.446
Within Groups	34.587	21	1.647		
Total	37.352	23			

Table 7.g Results of one-way ANOVA for bootstrapped means of responses to Factor 2: Self-reported Use of Head Movement and Impact of head angle

This result did not support the hypothesis that significant differences in the self-reported use of head movement and impact of head angle would exist between groups.

7.4 Self-Report Questionnaire: Content Analysis

The qualitative data was obtained from open-ended questions which were asked both before, and after head movement had been introduced as a topic of questioning. The qualitative results are presented first, after which the quantitative results are presented, beginning with data from responses given before head movement had been explicitly mentioned in the questionnaire.

7.4.1 Self-reported Auditory Sound Source Localisation Confidence

The content presented here was obtained from questions which were asked before head movement had been introduced as a topic in the questionnaire. Several issues of concern for the visually impaired arose in the qualitative data gathered, which could help to shed light upon factors that were detrimental to respondents self-reported confidence in sound localisation, as well as in navigating in the presence of sound under certain circumstances.

Perhaps unsurprisingly, both early- and late-blind participants mentioned "silent" vehicles as problematic, with electric cars and bicycles being frequently cited as harmful to the respondent's confidence when navigating busy environments. In fact, seven visually impaired participants voiced concerns over them, with mobility scooters mentioned by a further two. Their concerns were not limited to relatively quiet or silent vehicles however, and an interesting contrast emerged between the early- and late onset visually impaired groups: when asked about their confidence in estimating the distance to vehicles two members of the early-blind group reported that slow moving cars were problematic, in direct contrast to the late-blind group where only one respondent mentioned movement velocity, stating that "faster cars are harder". Four visually impaired respondents also mentioned that they had noticed the presence of interfering sound reduced their confidence, each referring specifically to different sources of said noise: sirens, helicopters, roadworks, and vehicles on roads adjacent to those of concern. In contrast, only one member of the sighted group mentioned interfering sounds reducing localisation confidence during the study, however two others specifically mentioned aircraft being hard to localise. In general, the responses from the sighted group were rather diffuse, only seven respondents mentioned certain sounds being harder to locate and, aside from the already mentioned aircraft, all were specific to the individual: phone alerts, burglar alarms, heavy goods vehicles, low-pitched sounds, and high-pitched sounds.

At this stage of the questionnaire participants were also asked to report on any sounds which they felt particularly confident when locating in terms of position, distance, or movement direction. Only three visually impaired participants responded affirmatively, two of which stated that crossing signal sounds were easy to locate, another stating confidence in locating human voices.

It is interesting that crossing signal "beeps" should be a point of confidence for some respondents, when emergency vehicle sirens are reported as detrimental to it when both stimuli occupy similar high frequency bands of sound. It seems reasonable to suggest that the crossing signal represents a fixed and stationary point, identifying a tool to assist navigation, whereas a siren represents a moving sound which is both loud enough to mask other sounds around it and is, by nature of design and cultural expectation, intended to alert people nearby to the danger of a fast moving vehicle which may not conform to the behaviour of other traffic around it (it may exceed the speed limit, pass traffic lights regardless of signals to stop, and continue to move when other vehicles become stationary). These factors may account for the stark contrast in the impact the two sounds have on visually impaired people's confidence when navigating, and when locating sounds.

7.4.2 Self-Reported use of Head Movement in Auditory Localisation

The content presented in this section was obtained from questions asked after the topic of head movement had been introduced in the questionnaire.

For determining the direction of sound source movement, five early-blind participants specifically mentioned that they employed a head turning strategy. Turning to face the sound was the most popular strategy, however one respondent stated that they turned one ear towards the sound and reported an intent to differentiate between reverberation and direct sound:

"I stop walking and turn my head, especially for sirens- they tend to echo very loudly so I want to tell the difference between echoes and the sound…I think if I focus one ear towards the sound I can pick up more information about it"

Although its impact upon auditory cues cannot be known from the response, this strategy of turning only one ear towards the perceived direct sound may be used to increase inter-aural level differences between direct and reverberant sounds by introducing the acoustical shadow of the head to separate the direct sound from nearby reflectors, a strategy which may help unmask sounds in a similar way to that observed in “cocktail party problem” type situations. Unlike the early-blind group, only one person from the late-blind group mentioned the use of head movement at this stage, reporting “careful turning towards” vehicles with high pitched engine notes, such as motor scooters. Three sighted participants mentioned head movement in the context of detecting sound source motion, in which two stated that they turned to face the sound, and one stated that they tended to turn one ear towards the sound.

Only two members of each visually impaired group mentioned head movement in conjunction with distance discrimination, and only one of these respondents (from the late-blind group) had not already mentioned using head movement in the context of determining the direction of sound source movement. This may mean that when directly asked about head movement later in the questionnaire (the results of which are included in *figure 7.e* and *7.f*) some visually impaired participants over estimated their use of head movement. For instance, ten respondents indicated that they used head movement at least some of the time to assist in sound source movement discrimination, three more than had previously mentioned head movement or angle before it was specifically described by the questionnaire. This result was not echoed by the sighted group, of which seven mentioned the use of head movement to facilitate distance perception before head movement had been introduced as a topic of questioning.

Finally, at the end of the questionnaire, respondents that had reported the use of head movement were asked to explain how they felt its use was helpful to them in the scenarios that had been described. Although only four participants felt able to answer this question, amongst the early-blind group two stated that they felt it helped to eliminate confusion between sounds that were of interest and those that were not, and one stated that they specifically tracked particular sound sources using head movement. In the late blind group one participant reiterated that head movement helped them to "hear more accurately", but was not able to say why. The sparsity of insightful comments here may be a sign of limitation in self-reporting, as well as issues in the construction of the question; when considering *how* head movement helps the participants may not understand the auditory processes underlying any reported benefit of head movement, should they have even considered it before questioning. Indeed, the answers that were offered describe better *what* the participants felt head movement achieved, rather than how it was achieved.

7.5 Discussion

The goal of this study was to provide a larger body of data regarding head movement and auditory perception than was provided by the field study discussed in *chapter 6.* For testing the hypotheses outlined in *chapter 7.1.3, the* lack of empirical data on this subject within literature made it difficult to estimate what data might be provided by a larger population, but the use of statistical bootstrapping has allowed for some assertions towards the original hypotheses to be made:

- Visually impaired individuals (VI) would self-report greater auditory localisation confidence than sighted individuals (S) (when responding to factor 1):
  $(H_1: \mu_{VI} < \mu_S)$

Although no statistically significant differences were found between groups with regards to this hypothesis, the mean confidence self-reported by visually impaired groups was actually lower than that of the sighted group, although all groups responded with some positive level of confidence ($\mu < 3$).

- Visually impaired individuals (VI) would self-report head angle and movement as a factor in auditory distance perception during navigation (number of VI responses to factor 1 at < 3 is greater than 40%[52] of VI total responses to factor 2)

[52] Chance level

The responses of the visually impaired groups indicated the use of head movement and awareness of the impact of head angle at levels above chance (66%).

- The self-reported use of head angle and movement to facilitate auditory perception in individuals with early onset sight-loss (E) will vary significantly from that of those with late-onset sight-loss (L), and sighted individuals (S) (when responding to factor 2):
  $(H_1: \mu_E \neq \ \mu_L \ AND \ \mu_S)$

There were no outstanding variances between participants with early and late-onset sight loss in regards to the self-reported use of head movement.

Moving beyond the tested hypotheses, it was also clear that there are multiple concerns for the visually impaired: electric or “silent” vehicles such as bicycles were often reported as a source of concern, as were sirens. It may seem surprising that sirens were a source of concern in terms of auditory localisation however, common siren sounds have previously been observed to be problematic in this regard [255]. Whilst this may not seem of direct concern to designers of assistive auditory VR, it is valuable information with regards to the design of symbolic auditory representations where avoiding similar types of sound design would be advisable. Reported sound sources or types which were difficult to localise for the sighted group were more varied and generalised, perhaps owing to their ability to visually verify the position of sound emitting objects, as was the case for the other qualitative components of the study. Qualitative results in the study only varied significantly between sighted and visually impaired groups for one question (question 5) in which the sighted reported greater confidence in determining whether vehicles were moving towards or away from them using the auditory channel.

Part IV: Acoustical Study

8. Analysis of Binaural Cues in Reverberant Space

This chapter describes a series of acoustical head-related transfer function measurements made in an enclosed reverberant environment, it explores interaural level differences, spectral properties, and interaural cross- correlation in direct and reverberant sound at varying sound source distances and head angles.

8.1 Rationale

The inclusion of reverberation in binaural virtual reality is known to mitigate the issue of in-head localisation/listening [162] [22] as well as to facilitate auditory distance perception [79]. In Bronkhorst's model for distance perception in rooms, absolute arrival times were considered the most significant in determining whether an arriving sound pressure wave would be considered as a direct or reflected sound by the human hearing system [86]. This model could predict perceived distances in listening tests with sighted participants, over distances up to 3m.

The present experiment aims to explore interaural cues, particularly interaural level differences, spectral cues, and cross-correlations over greater distances (up to ≈10m) and for varying head angles. This exploration of the impact of head angle could provide insight towards the need for accurately modelled reverberation in virtual environments under conditions where some form of head movement has been implemented.

For the sake of noise isolation, it was necessary to conduct the experiment described here indoors, which introduces factors such as ceiling reflections which would not be encountered in most outdoor environments. It is still of interest to see to what extent level differences exist between direct and reverberant sound at larger distances, and varying head angles.

8.2 Methodology

8.2.1 Measurement Apparatus

- Phonic PAA3 handheld audio analyser
- GRAS KEMAR 45BB-x Dummy head, torso, and KB5001 pinnae

- GRAS 40xx measurement microphone cartridges
- Brüel & Kjær Geodis omni-directional loudspeaker and amplifier
- Zoom H4N Portable digital audio recorder
- Apple MacBook Pro
- M-Audio FastTrack USB audio interface
- Time logarithmic sinusoidal sweep from 20Hz-20kHz (10 second duration, in .wav digital audio format)

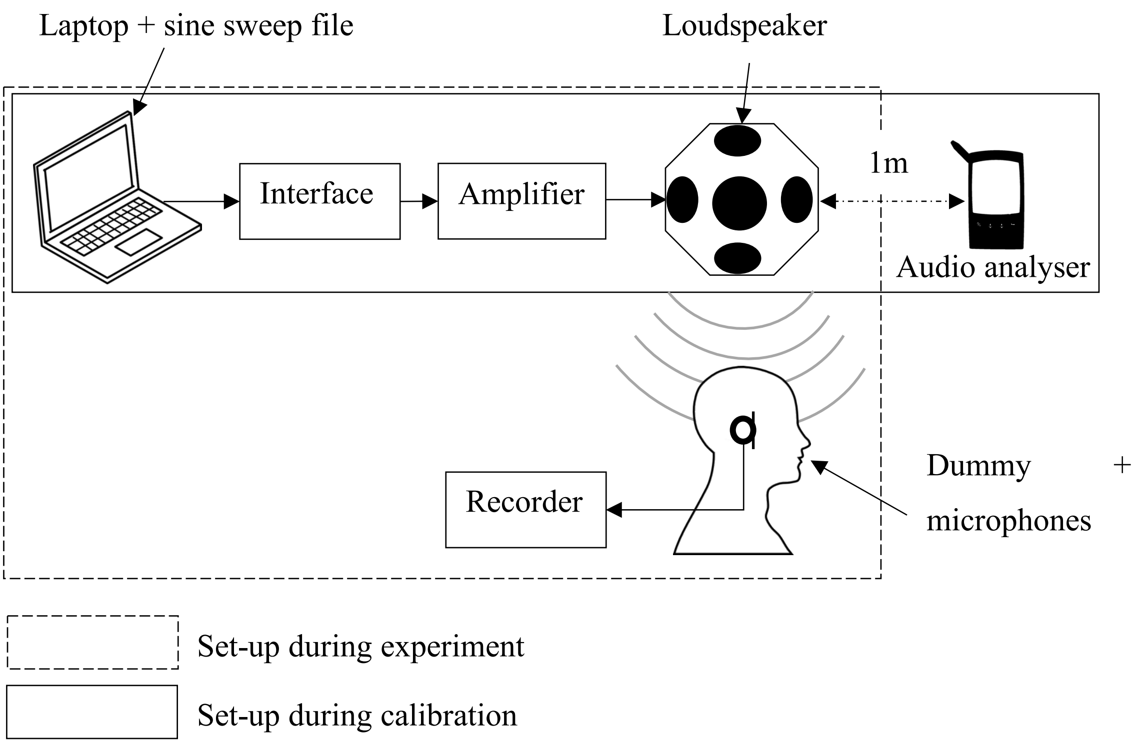


Figure 8.a Simplified diagram of both the calibration and experiment measurement equipment, showing audio/acoustical signal flow, and connection order of the test apparatus

8.2.1 Apparatus Calibration

Acoustical measurements were performed in a large room measuring 13.8m(L), 12.8m(W), and 9m(H) using a KEMAR dummy head and torso with two GRAS measurement microphones mounted internally at the ear canals for measurement recording, and a Brüel & Kjær Geodis omni-directional loudspeaker and amplifier as the sound source. The source audio was a logarithmic sine sweep from 20Hz-20KHz at 91dBSPL (at 1kHz) at 1m from the speaker.

Although cuboid in shape, the room was not a truly homogenous acoustical environment. The wall towards the right ear of the dummy head contained 2 pairs of wooden double doors, each pair measuring ≈3.4m(w) x 1.98m(h). Additionally, a set of un-removable fabric drapes running from the ceiling to the floor were in each side corner of the room towards the left ear of the dummy head. These drapes were drawn back to present the smallest surface area possible, resulting in an ≈2.5m(w) x 9m(h) portions of the left most wall being occluded at either corner. The RT_{60} of the room was measured at ≈2 seconds using the PAA3 audio analyser.

8.2.2 Experimental Measurement Positions

The dummy head was placed 3.5m from the rear wall of the room, and 6.54m from the side walls, such that each was equidistant to the position of the dummy. This position was maintained throughout the experimental measurements, whilst the omni-directional loudspeaker was placed initially at 1m from the dummy, and was moved back to 2m, 5m, 7m, and 10m. At each distance, a measurement of the sine sweep was recorded with the dummy head angled at 5º increments between 0º and 45º, and then at 15º increments between 45º and 90º.

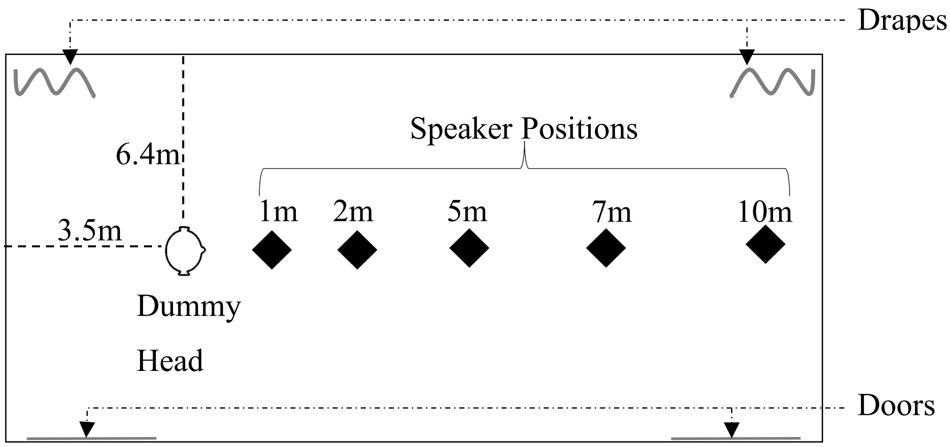


Figure 8.b the position of the dummy head, and sequential loudspeaker positions used during the acoustical measurements

8.2.1 Measurement Impulse Response Preparation

The recorded measurement sweeps were then convolved with an inverse of the original sweep signal, creating an impulse response measurement [256] of both the environment and the dummy head/torso for each of the head angles at each distance.

The impulse responses representing each angle and distance measured were separated into two time windows:

- The first window representing direct sound and reflections from the dummy head and body, covering the first 2 milliseconds of the impulse response; to include reflections from obstacles no greater than ≈68cm away from the ear canal, a radius which covers both the head, shoulder, and chest surfaces of the dummy [257]:

$$t = 2ms = \frac{(r = 68cm)}{(c = \ 343m/s)}$$

Where r is the distance from the ear canal opening of the dummy to the approximate position of the navel, and c is the speed of sound in room temperature (21ºC) air.

- The second window representing the remainder of the acoustical information regarding room acoustical characteristics (from 2ms to 2 seconds = RT60 of the experiment room).

8.3 Sound Pressure Levels for Direct and Reflected Sound

An analysis of direct to reflected sound pressure was conducted, in which reflected sound was further subdivided into early reflections (for reflections arriving between 2ms-50ms) and late reflections, or reverb tail (which was comprised of all sound arriving after 50ms). SPL values for each of the 3 windows were determined via RMS[53] averaging.

As would be expected, the direct sound shows a uniform reduction in SPL over distance, such that for each doubling of distance between the dummy and speaker the SPL is approximately halved. Although both early reflections and full reverberant SPL show a net reduction over distance, they do not conform to the same inverse proportionality as the direct sound. Furthermore, slight increases in SPL can be seen between measurements at 1m and 2m for early reflections.

A disparity was noted between early reflections and full reverberant SPLs measured between 5m and 10m, where the early reflection SPL shows a marginal increase, whereas the reverberant SPL increases <1dB between 5m and 7m, then decreases between 7m and 10m.

[53] Root-Mean-Square averaging, used to derive the time averaged amplitude of sound pressure (Errede, 2002)

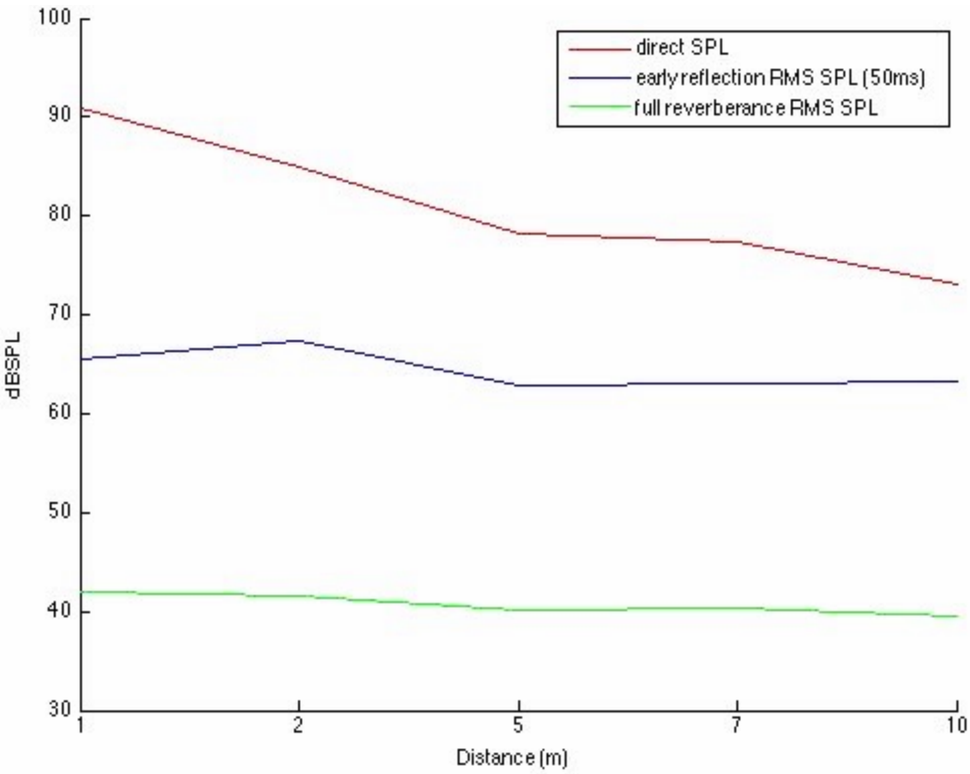


Figure 8.c A graph showing the measured direct (red), early reflection (blue), and reverberant (green) sound pressure levels after RMS averaging

It is expected that the summed SPL of all discrete reflections in a room will become roughly constant at any distance between the speaker and dummy head. This is because the SPL of reflected sound depends upon the geometrical configuration of the space within which the sound is held, and upon the acoustical properties of sound reflecting surfaces in the room. The results show a slight net decrease of ≈3dB between 1m and 10m measurement positions. This decrease is non-uniform, with both early reflections and total reverberant SPLs fluctuating in a way which may be accounted for by the presence of nodes and antinodes in acoustical standing waves at the point of measurement, which would (unlike average SPL levels) be dependent upon the position of the dummy head relative to the reflective surfaces and speaker in the room.

It is also possible that as the arrival angle of reflected sounds changes, because of the movement of the sound source, the arriving wave fronts may encounter acoustical impedance caused by the pinna, tragus, and head.

8.4 Interaural Level Differences in Direct and Reflected Sound

An analysis of interaural level differences (ILDs) in direct sound and early reflections was conducted. In this case, direct sound was any sound arriving before 2ms, and reflected sound was considered as any sound arriving thereafter. Once again, SPL levels for each window were obtained via RMS averaging.

In general, the ILDs in direct sound increased as the angle of the head increases. The ILDs in reflected sound by a lesser amount. In fact, ILDs in reflected sound are never greater than 10dB, even when the corresponding direct sound ILD is ≈18dB.

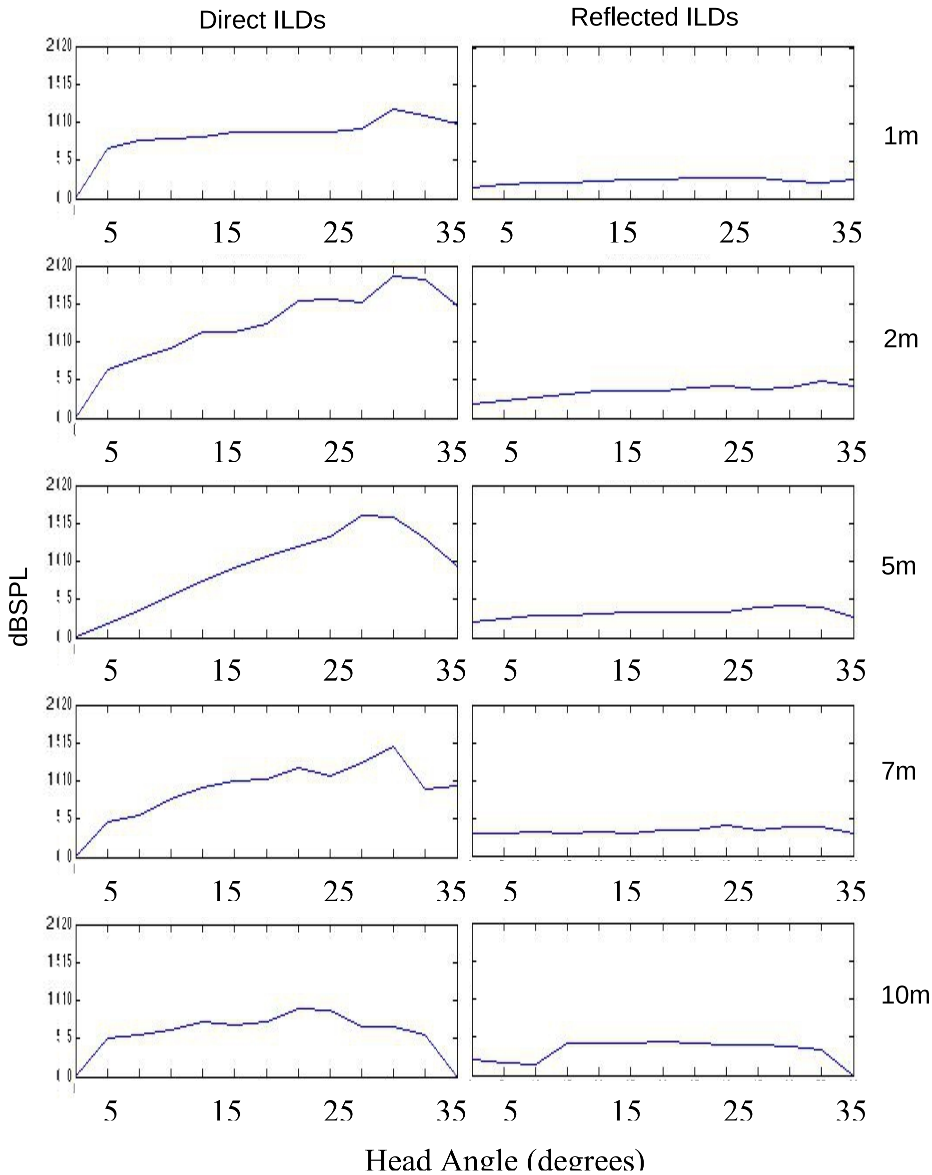


Figure 8.d Graphs showing ILDs in direct (left) and reflected (right) sound as a function of head angle and distance

The relatively low ILD at 1m, when compared to those at greater distances may be an anomaly. It would be expected that ILDs for sound sources closer to the listening position would be greater, as the distance between the ears represents a proportionally greater amount of the total distance to the sound source; according to inverse square law this should mean that energy loss (and therefore pressure decrease in the sound wave) would be measurably larger at the ear angled away from the sound source, although similar fluctuations in ILD measurements at different angles have been noted before, as sound in bounded spaces such as rooms does not strictly conform to the same pressure loss over distance as sounds in free/diffuse fields [258].

The disparity in ILDs between direct and reflected sounds can be accounted for by the directionality of the arriving sound wave fronts; reflected sounds arrive from points around the listening position, corresponding to the positions of the reflective surfaces (walls, floor, and ceiling) of the room, whereas the direct sound arrives from the direction of the sound source alone.

As predicted by the relatively small deviation of SPL for reflected sound at all distances measured, the ILDs for reflected sound are consistently below 5dB, regardless of head angle or distance between speaker and dummy head.

The relatively low ILDs in direct sound at 1m cannot be easily accounted for. In general, ILDs from distances of 2m or greater proceed as expected, that is they decrease as distance to source increases- this would indicate a possible error in the experimental procedure when the ILDs at 1m meter were measured, or when the resulting audio files were processed for analysis.

8.5 Interaural Level Differences over Frequency and Distance

The final analysis of ILDs over both distance and frequency was conducted for both direct and reflected sound. The impulse responses were divided into 8 frequency bands with centre frequencies of 250Hz, 500Hz, 1000Hz, 2000Hz, 3000Hz, 4000Hz, 5000Hz, and 6000Hz using 3rd order Butterworth bandpass filters. As with the previous analyses SPL for direct and reflected sound was determined via RMS averaging, with direct sound considered as any sound arriving before 2ms and reflected sound being any arriving thereafter.

At an azimuth angle of 0° ILDs in reflected sound are greater than those of direct sound. With the sound source located <2m from frequencies from 250Hz-6000Hz is consistently lower than that of the direct sound (except for the 0° azimuth measurement). As distance increases the difference between mean ILDs reduces, particularly for angles >30°.

At distances >5m the mean ILD of reflected sound is greater than that of direct sound when the sound is located at 5° of azimuth.

Maximum ILDs in direct sound are between 24dBSPL and 26dBSPL and are found at angles >30° in all measured distances between the sound source and dummy head. Maximum ILDs in reflected sound are between 20dBSPL and 23dBSPL and are generally found at angles of 20° to 35°, although they are inconsistent over the range of distances measured.

The only consistent azimuth angle at which mean reflected ILDs are greater than direct ILDs is 0° reinforces the assertion that ILDs in reflected sound depend upon factors beyond the angle of the sound source relative to the head, they must also depend upon the geometric configuration and acoustical properties of the space within which the sound is contained. Considering that the mean ILD for reflected sounds is never greater than 10dBSPLat any angle or distance measured, even when corresponding mean ILDs for direct sound could be as great as ≈17dBSPL this assertion is further reinforced.

8.6 Spectral Analysis of Impulse Response Measurements

Spectrograms of the measurements made during the experiment, at different distances and angles, for both ears were computed, and spectral differences in measurements over both head angle and sound source distance were computed as spectrograms (the methodology for which is described in *8.6.1).*

One of the most prominent features visible in the measurement spectrograms was apparent spectral notches caused by the pinna [259], below 15kHz. These notches were present in both direct sound, early reflections, and reverberation across all distances. They could also be consistently observed in either ear at any head angle.

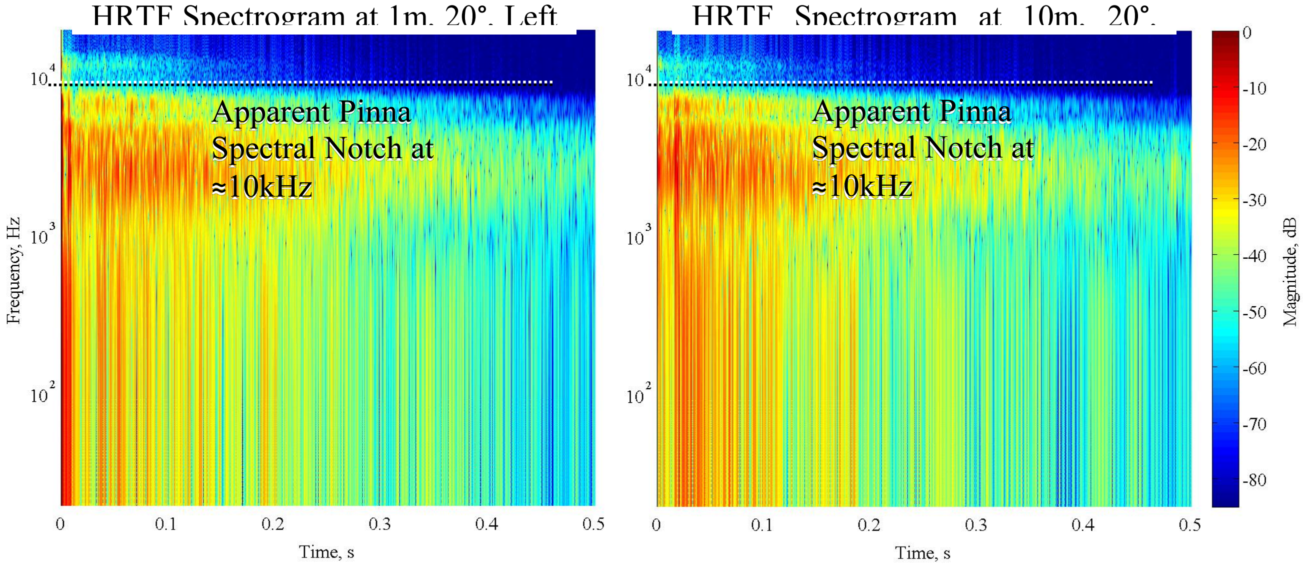


Figure 8.e Spectrogram of measurements taken at 10m from the sound source for both left and right ears at 60° azimuth, with the pinna spectral notch illustrated

These notches are most significant for the perception of sound source elevation, with the frequency at which they are present having an inverse relationship with the angle of elevation [260]. Also, visible in these spectrograms is the difference in early arriving reflections and direct sound. In the above figure, the earliest arriving sound waves are noticeably lower in magnitude in the right ear (furthest from the sound source) owing to early reflections created by boundaries between the listening position and sound source which arrive at the closest ear first, encountering less impedance from the listener's head. At closer distances, although this attenuation is still present, it is less severe, particularly below ≈1kHz. In the left ear (closest to the sound source) this level difference is present, although less pronounced.

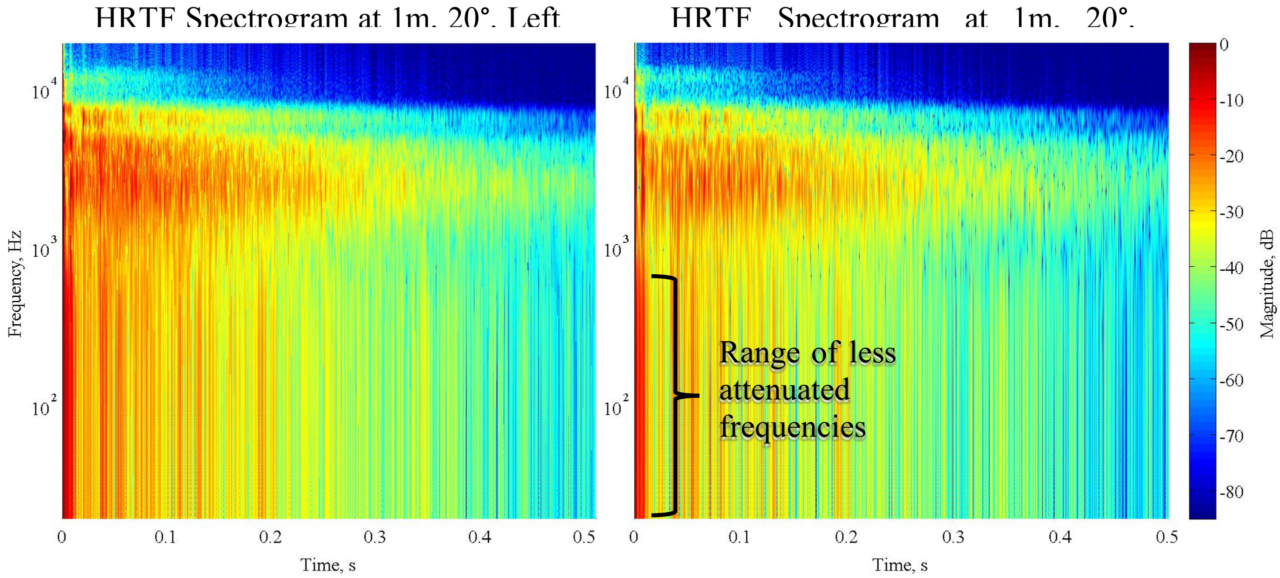


Figure 8.f Spectrogram of measurements taken at 1m from the sound source for both left and right ears at 20° azimuth, with range of less attenuated direct and early arriving reflections illustrated

This range of frequencies encountered the greatest attenuation. At 10m, with the sound source at 20° azimuth or more, up to >10dB of attenuation was measured at the right ear in early arriving reflections above 10kHz when compared to the left ear.

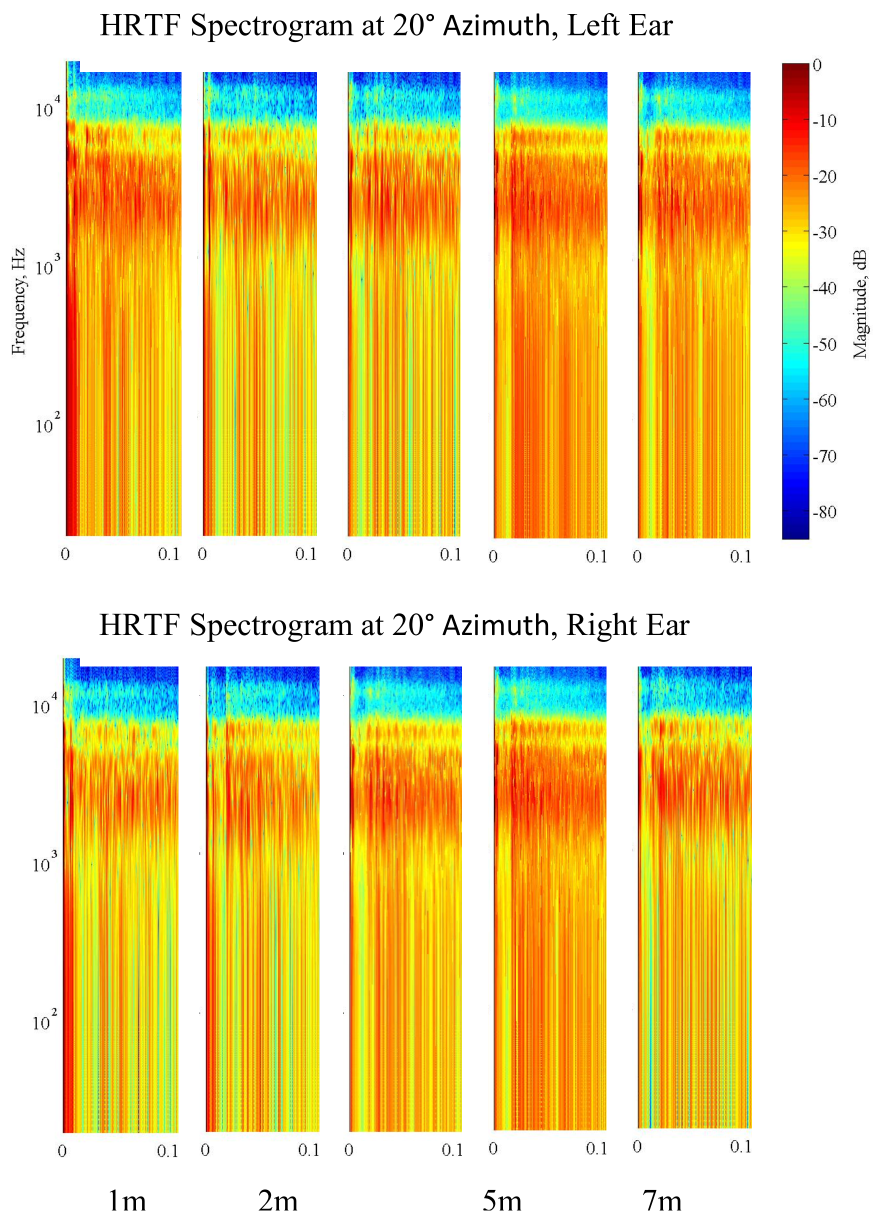


Figure 8.g Spectrograms comparing the amplitude of sounds arriving within .1 seconds of the direct sound for left (top) and right (bottom) ears at 20° azimuth at distances between 1m and 10m

As the comparison above shows, sound pressure arriving at the right ear is generally lower than that arriving at the left ear (for angles from 20° azimuth), however the greatest differences between left and right ears all appear below 6kHz and within the first ≈25ms of arriving sound. This change is not only a result of ILDs but also differences in the position and/or intensity of early reflections, which vary according to the relative position of reflecting boundaries as the angle of incidence and reflection is altered by the movement of the sound source.

8.7 Interaural Cross Correlations

A final analysis of the cross correlations between ears at each distance and angle was conducted.

8.8 Interaural Cross Correlations Over Distance and Angle

Interaural Cross Correlation Coefficients (IACCs) were calculated for each series of measurements. The IACCs were used to determine maximal correlation levels between wave fronts arriving at each ear, as well as interaural time differences but the arriving waves.

First, IACCs for direct sound were computed by isolating the first 2ms of the measured impulse responses (shown in *figure 5.h*).

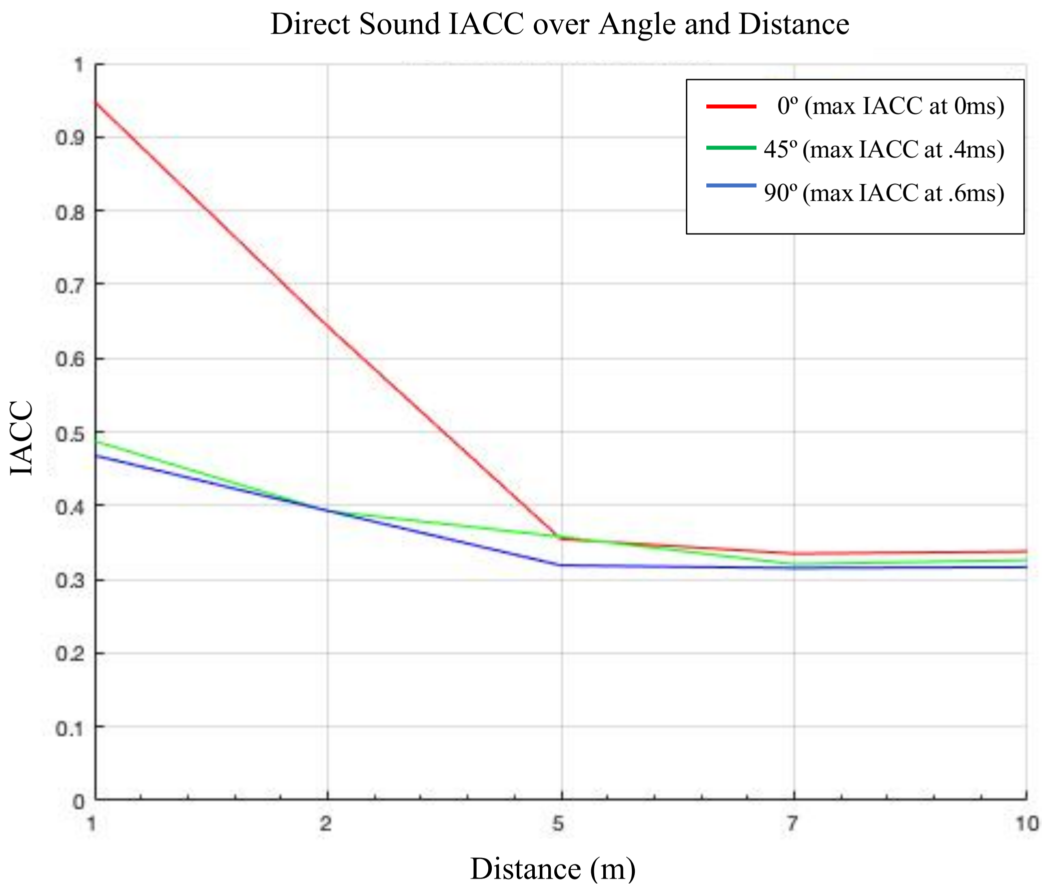


Figure 8.h Plot of IACC for direct sound over angle and distance, showing IACCS for 0º (red), 45º (green) and 90º (blue)

Here, the highest IACCs were measured at 1m, and for measurements where the source was at 0º azimuth. IACCs decreased greatly as a function of both angle and distance, with the greatest reduction shown between 1m and 5m. At 5m, 0º and 45º IACCs are almost identical. As would be expected, the ITD varies consistently with the dimensions of the head as a function of the angle of the head relative to the sound source. This indicates that the direct sound impulse is indeed the most significant inter-aurally correlated waves in the measurement.

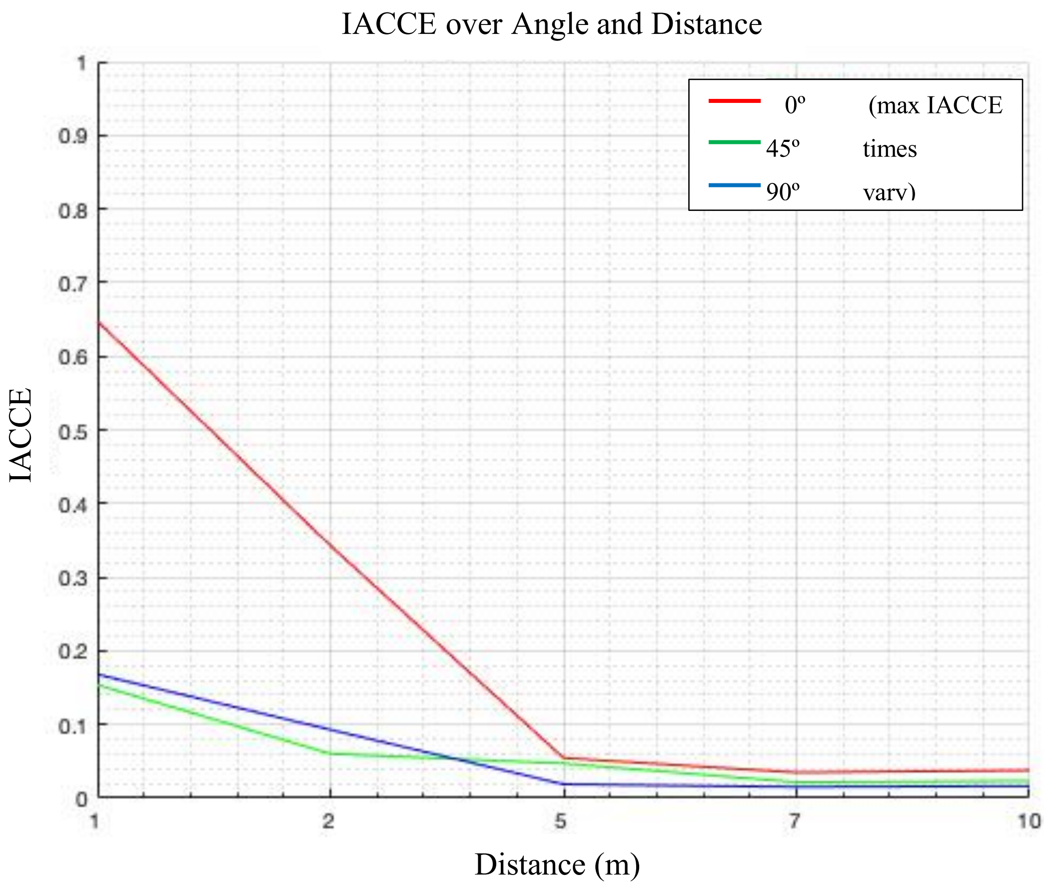


Figure 8.i Plot of IACCE over angle and distance, showing IACCs for 0º (red), 45º (green) and 90º (blue)

Next IACCs were computed for the early (IACCE[54]) portion of the impulse response measurements (shown in *figure 5.i*). Here, the highest IACCs were measured at 1m, and for measurements where the source was at 0º azimuth. IACCs decreased greatly as a

[54] IACC calculated over the first 80ms of a measured signal, excluding direct energy

function of both angle and distance, with the greatest reduction shown between 1m and 5m. The lag time of maximal correlation time in isolated early reflections with the sound source placed at 45º azimuth rose as a function of distance between the dummy head and sound source as follows: 1m ≈ .41ms, 2m ≈ 1.92ms, 5m ≈ 7.81ms, 7m ≈ 13.13ms, 10m ≈ 30.85ms. This increasing correlation time are likely a product of the dimensions of the space in which the measurements were made, where highly correlating early reflections from surfaces closest to the speaker and dummy head arrive at varying times as their propagation path between speaker and dummy is lengthened.

Finally, IACCLs[55] were calculated (shown in *figure 5.j*).

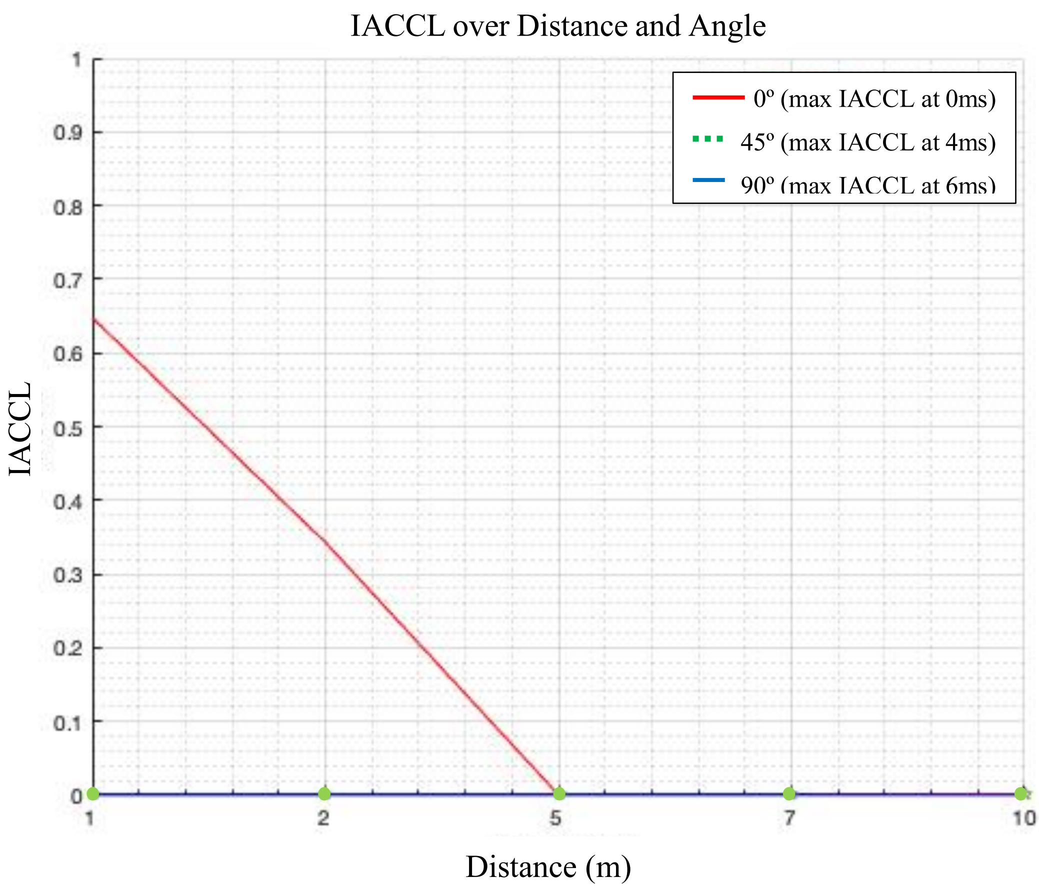


Figure 8.j Plot of IACCL over angle and distance, showing IACCs for 0º (red), 45º (green) and 90º (blue) and maximal interaural correlation times within the late reverberation window

---

[55] IACC calculated over the measurement period, excluding direct and early reflected energy

Here correlations were observed for sounds located at 0º azimuth, at distances <5m, outside of which no distinct correlations were found.

8.9 General Discussion

Given that, in reverberant environments, just noticeable differences[56] arising as a result of ILDs have been demonstrated to be a more robust cue for sound source localisation than those arising because of ITD [261], the ILD differences detected in impulse response measurements may be a more viable cue than the ITDs. The level differences in reverberant energy measured here were lower than those found in direct sound, and varied less as a function of head angle.

The highest correlations in early reflections for sounds emitted within 2m of the dummy head were found within 2ms. For sounds arriving after this time frame, correlation approached zero at all angles outside of 0° azimuth. The data measured here would support the notion that accurately modelled early reflections would provide the most salient psychoacoustical cues, and a generalised algorithmic late-reverberation tail could be employed, if it provided appropriate direct-to-reverberant energy ratios for the space being virtually modelled. Whilst this experiment demonstrated the phenomenon within a large reverberant room, in a typical urban environment, the presence of two reflecting surfaces (such as the floor, and one nearby building or wall) could generate early reflections, although the RT_{60} would likely be lower than in a fully enclosed environment [262].

---

[56] Differences which are considered perceptually detectable by humans

Part V: General Discussion & Conclusions

9. General Discussion

This section discusses the scope of the three studies presented and compares the key findings in each with information drawn from the literature review with a view to commenting upon the use and role of head movement in auditory perception of the visually impaired. Based upon this, suggestions are made for designers of virtual reality to consider the use cases of their technology and the end user's requirements, with a view towards reducing where appropriate the complexity and cost of assistive virtual reality technology. The original contributions are summarised and finally, some future avenues of study expanding upon the work presented are suggested along with the contextual framework within which they sit.

9.1 Scope of the Studies Presented

It should be noted that the field study reported in *chapter 6* was conducted over a very small sample of individuals. Although the sample was well proportioned in terms of participants with early onset and late onset sight-loss, this sample alone is too small to make a strong case for the use of head movement in auditory perception amongst the visually impaired. The self-report questionnaire study reported in *chapter 7* contains a larger sample of 26 (13 of which were visually impaired individuals), and utilised statistical bootstrapping upon two factors identified within the questionnaire to make inferences regarding the larger visually impaired population. These replicated sample results were used to minimise the risk of type I errors[57] in the one-way ANOVA tests which were used when analysing the results. Although steps were taken to minimise biases in results introduced by social desirability, and by prior learning or expectations concerning navigation, such biases cannot be truly considered excluded.

The outstanding limitation of the questionnaire study was its use of self-report methods. As discussed in *chapter 7*, whilst this method may be used to assess people's behaviours (particularly those which they assign personal importance to) it cannot offer reliable access to an understanding of the deeper cognitive processes which may underlie that behaviour. Nonetheless they offered a viable source of insight into the self-reported navigational strategies of the visually impaired, where the application of stronger observational methods described in *chapter 6* had been weakened by small sample sizes. Qualitative data regarding the visually

[57] The erroneous rejection of a null hypothesis

impaired was also of value, as certain trends in the self-reported sounds and scenarios of interest or concern for the visually impaired emerged.

The acoustical study presented in *chapter 8* provided an acoustical analysis of available auditory cues in a large, reverberant environment. Although this environment is removed from the typical outdoor areas that were the topic of the previous studies, it did offer some insight into the changes in acoustical cues which are introduced in conditions where the distance of the sound source, and the angle of the head relative to the position of the sounds source, change. Whilst they remained subjectively unexplored, a number of potentially viable related cues were observed.

9.2 The Use of Head Movement by the Visually Impaired

Head movement was both measured, and self-reported, in the visually impaired during navigation tasks. This included both those participants with early and late-onset sight loss. Whilst it is possible that the self-reported use of head movement was a result of a perceived expectation to report such, the combination of measured behaviour and the fact that some participants reported the use of head movement before it was specifically addressed in the questionnaire strengthen the case for this being a valid result. Although little quantitative variance beyond head movement (yaw) velocity was found between visually impaired and sighted individuals (meaning that the initial hypothesis that there would be quantitative variance between early-onset, and late-onset and sighted groups was unsubstantiated) there were qualitative differences in the use of head movement measured in the observational study (detailed in *chapter 6*), as well as some qualitative differences in the self-reported strategies (detailed in *chapter 7*). As may have been expected, the sighted appeared to use head movement as a tool for visually verifying objects such as crossing lights, and trip hazards, whilst the visually impaired responded more to sound sources.

The visually impaired also discussed the hazards posed by electric cars, bicycles, and other vehicles which produce relatively little sound, where the sighted reported auditory localisation issues in more generalised terms. These qualitative differences in the self-report responses seem likely to have arisen because of the importance and priorities different groups placed upon sound during navigation: “silent” vehicles would pose a risk to the visually impaired as, in the absence of sound, their presence would be hard to identify. The sighted, possessing the ability to visually verify the location of such vehicles, would likely feel less threatened by them and therefore place less importance upon them in the context of the navigational studies presented here. This may have also been a factor in the observational study- the visually impaired tended to take longer to cross pedestrianised roads than the sighted. The waiting strategy they displayed at this crossing may have been an effort to determine if any silent vehicles or pedestrians were nearby, as they had been informed that this crossing was closed to road vehicles. Indeed, during the self-report study, a number of visually impaired respondents mentioned standing still as part of their sound localisation strategies. These self-reported strategies also included some element of head movement- most common was turning to face the apparent source of a sound, although turning one ear towards the sound was also described by one participant. Additionally, standing still would reduce self-noise, which may be helpful when attempting to locate other pedestrians walking nearby. In the final questions of the self-report study, the visually impaired gave the strongest indication of why they believed head movement to be useful: it was felt that it was particularly helpful in determining a change in distance/range of sound sources.

As described by Blum, *et al* [173] the implementation of head tracking rather than joystick control offered no improvement to visually impaired user's abilities to locate sound targets in virtual reality- a fact that supported Wightman's earlier [150] finding that emulated head movments, whether controled by head tracking or not, offered reduced localisation errors so long as the user had control of said movment. In these cases the sound targets were static, only the user moved within the environment. Given the responses to the self-report study, it is possible that results for moving sound targets may be improved by head tracking, although it is also likely that some improvement would be offered by manual (such as joystick) control of emulated head movmeent. In this case the observational study detailed in *chapter 6* does provide some empirical data that would support the implementation of head movement without the need for head tracking: visually impaired individuals demonstrated typical head yawing velocities of up to ≈9.42 radians, or 539 º per second, particularly at road crossings. Pitch and roll movements were much slower than this- typically ≈.12 radians, or 7º (for pitch) and ≈1.57 radians, 90º per second (for roll). It is also of interest to note that rolling movements sometimes occurred at the maximum extent of a yawing motion, a complex motion which may need further consideration when implementing some manual control for head movements as none of the participants in the self-report study (*chapter 7*) indicated any awareness of this type of rolling movement- if it is an unconscious but useful movement then it should be quantified and implemented in some way.

Regarding the function of head movment in audition, Bronkhorst's model of auditory distance perception [84] predicted the importance of direct to reverberant energy ratios on distance perception. The experiment presented in *chapter 8* of this thesis showed that for distances of up to 10m the variance in ILD as a function of head angle was much greater for direct than reverberant energy ratios. Although authors including Begault, *et al* [162] have shown that reverberation can confound precise angular localisation there is a clear precedent for the use of reverberant energy in distance estimation, and although the results given here do not indicate precisely *how* head movement may facilitate this, they certainly do not preclude it.

The spectral analysis of direct and reverberant energy presented in *chapter 8* reveals some interesting phenomenon such as the obvious presence of pinna spectral notches throughout sound reflections at all distances, and the prominence of lower-mid frequency differences in the earliest arriving reflections. It is beyond the scope of the study in question to determine how significant spectral content is in distance estimation, particularly in unfamiliar environments. One thing that is made clear however, is that a complex of spectral differences exists in both direct and early reflected sounds. These may at least provide a case for the use of accurately modelled reverberation in virtual acoustic environments, rather than algorithmic reverb which may not provide sufficient differences in early sound reflections- particularly in the context of studying auditory distance perception in the presence of sound reflecting surfaces/boundaries.

The comparisons of IACC as a function of distance and angle showed that higher interaural correlations were found within the early reflection window of the reverberation, but in the late tail such correlations were greatly reduced, and in fact not truly present when the head angle relative to the position of the source was increased, indicating that accurate early reflection modelling may be sufficient to recreate reverberant auditory cues in a virtual environment, even under conditions of head movement.

9.3 Towards a Guideline for Developers of VR for the Visually Impaired

Head movement cannot be excluded from the design of virtual reality systems for the visually impaired; it has been shown to be used by them in at least some aspects of auditory distance perception. It is also of use for reducing common localisation errors in binaural audio such as front-to-back confusions and in head listening, however these issues have also been shown to be reducible by other means. Begault, *et al* [162] notably showed that reverberation reduced in head listening, and given reverberant energy's function in distance perception, it should certainly be considered a fundamental aspect of auditory; there are cases in which simulated reverberance is unnecessary, for instance if the goal is simply to reduce the phenomenon of inside head listening then algorithmic reverb, not approximating a real environment may be used (as proposed by Begault [7]). In situations where the boundaries of the environment itself are important, or where distance perception is needed, simulated environmental reverb with accurate portrayal of direct to reverberant energy ratios will always be desirable. Wersenyi [168] has shown that movements of a sound source itself can reduce both in head listening and front-to-back confusion to some extent and, more significantly, Wightman [150] found that any movement under a users control was sufficient to also reduce them; a case is made here for adopting these methods over that of tracked head movement as the attendant hardware required for them is more readily available than that of motion trackers.

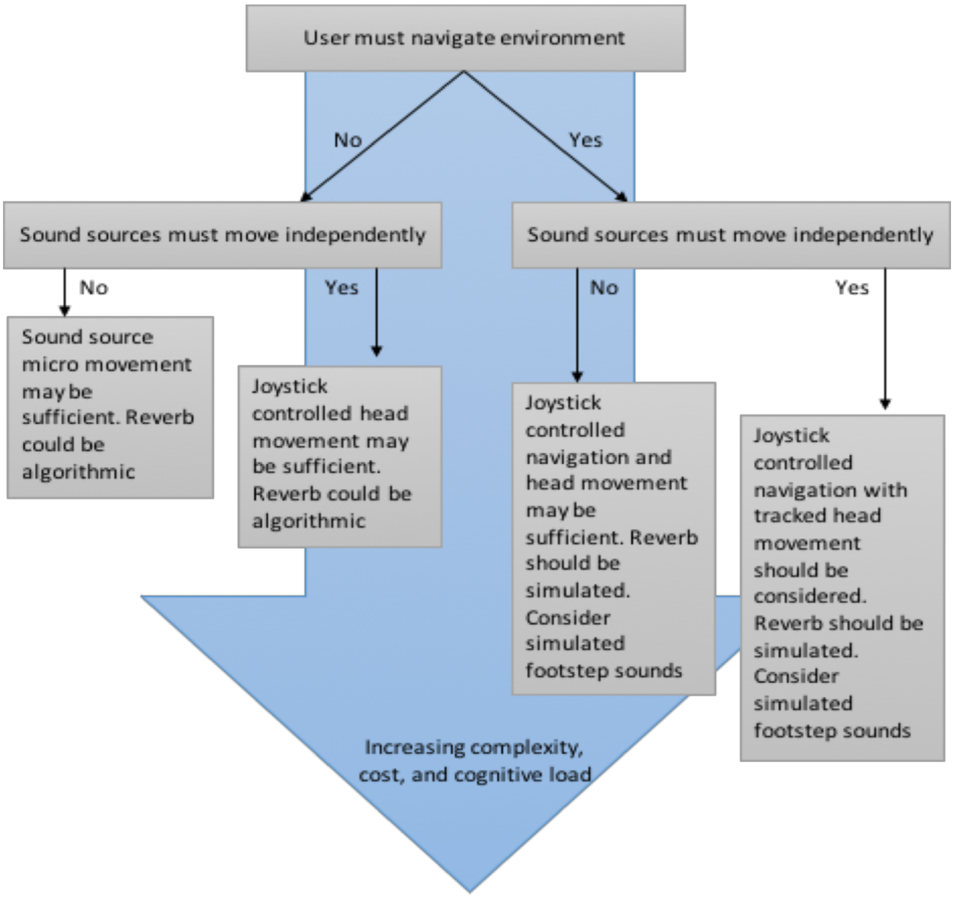


Figure 9.a Decision tree for design of assistive VR systems based upon the need for user interaction and auditory localisation

Returning to the topic of considering the end user, and use case, of a system when designing it some suggestions can be made based upon both the literature discussed, and the primary research, presented in this thesis:

9.3.1 Example use case Scenarios for the Proposed Decision Tree

Presented here are some short examples of systems which have arisen in literature, accompanied by a brief outline of the design decisions as per the proposed decision tree, with some justification for feature inclusion:

I. An auditory display for rendering computer desktops/software interfaces:

This is the simplest application presented here; there is no need for the user to be presented with the illusion of exploring a space, aside from cursor movement there no need for a great deal of independent movement of sound sources, in fact, sound sources are designed as sonification to give an auditory representation of a visual icon on a computer screen.

Here the user requires no control input relevant to the virtual reality simulation, their interactions are directly with the interface being sonified, for this reason the use of randomly generated micro movements of sound sources such as those described by Wersenyi [168] present one solution to the common localisation ambiguities encountered in auditory virtual reality. Algorithmic reverb can also be included to reduce inside-head-listening. In this case, as there is no need for sound sources to bear any resemblance to real world sound sources, the sonification process can also be used to reduce factors that confound auditory localisation, for instance avoiding predominantly low frequency sounds in favour of sounds with significant energy between ≈700Hz and 1kHz, thus ensuring that level differences, as well as time differences are detectable [263].

II. A system for environmental map learning:

These types of applications have been focused on learning the configuration of a given space, with a view to allowing the user to learn information about a real-world locale for the purposes of enabling way finding [19]. Here the user must be able to freely navigate the virtual environment, however sound sources tend not to be independently mobile as the goal is to learn architectural configurations.

A system such as this could be controlled by joystick input for both exploration and head movement, allowing the user the control which plays a major factor in reducing the common localisation ambigutities in auditory virtual reality. Sound sources representing way points (for example a fire escape) can be represented with sonnification, as described previously. The designer may feel the need to include some sounds which represent a real world sound source, for instance outdoor ambient sounds such as wind and birdsong to indicate that the user is outdoors, or traffic sounds to indicate proximity to a street or exit proximal to a street. In these cases micro movements below the threshold of detection may also be employed to improve localisation accuracy.

The inclusion of simulated footsteps will provide a sound source for echolocation, and may also provide cues for movement rates and distances, though the appropriate stride distance and step rate remain untested for emulated walking in such an application.

Reverberation (particularly in the early reflection phase) should be acoustically simulated as to allow the user to detect the presence of sound reflecting objects such as walls. Accurate direct to reverberant energy ratios throughout the RT_{60} would allow for the best potential sound source distance estimation, however in environments with reverberant radii greater than 5m the reverberation tail may be approximated algorithmically with inter-aural cross-correlations approaching zero.

III. A system for learning safe navigation techniques:

These types of systems necessarily have high potential complexity, they may require the user to not only freely explore an environment, but to also be presented with independently moving sound sources in perhaps the most complex sound fields. An example application of this type of system would be a road crossing simulation, intended for the user to explore safe crossing techniques.

In this case joystick control for environmental exploration and head movement is perhaps the simplest method of control that would be acceptable. Head tracking would allow for a more natural experience, which could be quickly translated into head movement behaviour in the real world and vice-versa. Sound sources should mimic their real-world counterparts as closely as possible; it would be imprudent to design sounds that deviated from actuality as the goal is for the user to successfully navigate dangerous, real environments. Reverberation should be acoustically simulated to provide the most accurate distance cues possible. In outdoor environments

A further expansion (beyond those listed in the decision tree in *figure 9.a* would be the inclusion of individualised HRTFs. Although this would add great complexity and expense to a system owing to the need for the measurements to be expertly made, it would ensure the most accurate localisation experience for the end user.

The three systems presented briefly here represent a range of system functions and complexities. It should be noted that the most complex system (described as “A system for learning safe navigation techniques”) would in fact be compatible with the functions and design of the less complex systems; the less complex systems are only compatible with those of lower complexity. This is because many of the mechanisms employed by the lower complexity systems are achievable without the need for specialised hardware: sound source, micro movements, and algorithmic reverberation (where no measurements or predictions of detailed real-world acoustics are necessary) can be achieved through software methods alone. Individualised HRTFs could be applied to any of these systems, were the improvements they offered deemed necessary. A secondary feature of lower complexity systems is potentially lower system latencies, although the accuracy of active localisation tasks may not be affected by latencies of up to 250ms thanks to behavioural strategies adopted by users, maximum latencies of ≈30ms or less should be the goal for systems designed to emulate real world scenarios for safety training, where behavioural strategies should ideally be focused on learning safe navigational techniques rather than compensating for system deficiencies. As described in *chapter 4*, the most appropriate method for head related impulse response and transfer function measurement for any binaural application would be the logarithmic sine sweep, due to its robustness to technical issues such as signal distortion, and its ability to capture the spectral cues associated with the (as described by [55]).

10 Conclusions

10.1. Final Remarks on the Design Guideline

This model expands upon that offered by Begault [9] in that it expands upon the concept of system function vs. cost by adding a decision framework for designers of assistive, auditory virtual reality. The core topic of the primary research offered in this thesis- the use of head movement in auditory perception and navigation by the visually impaired- as well as secondary research into the field of human sound localisation have been considered in the design of this guideline and, although brief, it is conducive to the goal of reducing the complexity and cost of assistive systems; reducing the required attendant hardware for a system to that which is more commonly, and cheaply, available to the consumer allows a larger group of users to enjoy the benefits of such a system.

10.2 Contributions

Primary research into the use of head movement in auditory perception by the visually impaired has been presented here. Although many studies have explored the role that head movement may play in auditory perception, this research has contributed some way towards answering the question of whether visually impaired persons naturally use head movement to facilitate audition- this is perhaps the first time such a question has been posed, or an answer been approached.

A guideline based upon the primary and secondary research of this text has been created with the goal of aiding designers of assistive auditory virtual reality to reduce the complexity and cost of systems by considering what is functionally necessary for the end user and the intended functions of the system.

10.3 Future Work

Following the literature review and studies presented here, as well as the proposed model/decision tree for virtual reality system design, several unexplored or new directions arise. Although the global goal of bringing viable assistive virtual reality technologies for the visually impaired to market is the most obvious continuation of this work, some intermediate steps are worthy of consideration prior to a finalised system being presented:

I. An exploration of the role of the effect of simulated footstep sounds on users perceived movement speed and displacement distance in explorable virtual environments.

II. Further investigation of the impact of simulated early reflections with approximated reverberant decay tails in reverberation in echolocation and environmental cognitive map forming in virtual environments.

III. A study of head movement and reverberant distance cues comparing accurate acoustic cue modifications with control modifications to demonstrate the effectiveness of angle and movement upon distance perception

IV. An expanded study of typical head movement axes and velocities, sufficient to inform the design and instruction of manual head movement control systems

Additionally, when advancing designs for commercial virtual reality software and representations of environments, assessments of the most perceptual significant sound features required for the intended purpose of the environment should be considered, although this work does not arise directly from that which is presented here but form the necessities of system simplification discussed and demonstrated in the literature.

Appendices

Appendix One: Sensor Toolkit Firmware in Pure Data

```
/****************************************************************************
****************
* Test Sketch for Razor AHRS v1.4.1
* 9 Degree of Measurement Attitude and Heading Reference System
* for Sparkfun "9DOF Razor IMU" and "9DOF Sensor Stick"
*
* Released under GNU GPL (General Public License) v3.0
* Copyright (C) 2013 Peter Bartz
* Copyright (C) 2011-2012 Quality & Usability Lab, Deutsche Telekom Laboratories, TU Berlin
* Written by Peter Bartz (peter-bartz@gmx.de)
*
* Modified by Chris Feakes 2013-2014(cfeakes@dmu.ac.uk or cafeakes@gmail.com)
*
* Infos, updates, bug reports and feedback:
*     https://github.com/ptrbrtz/razor-9dof-ahrs
****************************************************************************
***************/

/*
  NOTE: There seems to be a bug with the serial library in the latest Processing
  versions 1.5 and 1.5.1: "WARNING: RXTX Version mismatch ...". The previous version
  1.2.1 works fine and is still available on the web.
*/

import processing.opengl.*;
import processing.serial.*;
import processing.video.*;
//import ddf.minim.*;
```

```
// IF THE SKETCH CRASHES OR HANGS ON STARTUP, MAKE SURE YOU ARE USING THE RIGHT SERIAL PORT:
// 1. Have a look at the Processing console output of this sketch.
// 2. Look for the serial port list and find the port you need (it's the same as in Arduino).
// 3. Set your port number here:
final static int SERIAL_PORT_NUM = 2;//for BT use: /dev/cu.RN42-1E0C-SPP pword: 1234
// 4. Try again.
final static String OUTPUT_TXT_FILE_NAME = "SUB B05 TEST COORDS";//Name txt file

final static int SERIAL_PORT_BAUD_RATE = 57600;

final static int VISUAL_COORDS = 1;//if 1 coord rendering on, if 0 rendering off

//Camera data recording addition by CF//

final static int CAMERA_PORT_NUM = 16;//select camera port number 30fps

final static String OUTPUT_VIDEO_FILE_NAME = "SUB B05 TEST VIDEO";//Name video frames

//Initialise video record on/off variable addition by CF//
int REC_VIDEO = 0;
//Initialise coord variables
float yaw = 0.0f;
float pitch = 0.0f;
float roll = 0.0f;
float yawOffset = 0.0f;
//Timer addition by CF//
int time = 0;


PFont font;
```

```
Serial serial;

boolean synched = false;

void drawArrow(float headWidthFactor, float headLengthFactor) {
  float headWidth = headWidthFactor * 200.0f;
  float headLength = headLengthFactor * 200.0f;

  pushMatrix();

      // Draw base
      translate(0, 0, -100);
      box(100, 100, 200);

      // Draw pointer
      translate(-headWidth/2, -50, -100);
      beginShape(QUAD_STRIP);
        vertex(0, 0 ,0);
        vertex(0, 100, 0);
        vertex(headWidth, 0 ,0);
        vertex(headWidth, 100, 0);
        vertex(headWidth/2, 0, -headLength);
        vertex(headWidth/2, 100, -headLength);
        vertex(0, 0 ,0);
        vertex(0, 100, 0);
      endShape();
      beginShape(TRIANGLES);
        vertex(0, 0, 0);
        vertex(headWidth, 0, 0);
        vertex(headWidth/2, 0, -headLength);
        vertex(0, 100, 0);
        vertex(headWidth, 100, 0);
        vertex(headWidth/2, 100, -headLength);
```

```
      endShape();

  popMatrix();
}

void drawBoard() {
  if (VISUAL_COORDS == 1)
  {
    pushMatrix();
  ///The roll and yaw offset is new, does it work?
    rotateY(-radians(yaw - yawOffset));
    rotateX(-radians(pitch));
    rotateZ(radians(roll));

    // Board body
    fill(255, 0, 0);
    box(250, 20, 400);

    // Forward-arrow
    pushMatrix();
    translate(0, 0, -200);
    scale(0.5f, 0.2f, 0.25f);
    fill(0, 255, 0);
    drawArrow(1.0f, 2.0f);
    popMatrix();

    popMatrix();
  }
}

// Skip incoming serial stream data until token is found
boolean readToken(Serial serial, String token) {
  // Wait until enough bytes are available
```

```
  if (serial.available() < token.length())
    return false;

  // Check if incoming bytes match token
  for (int i = 0; i < token.length(); i++) {
    if (serial.read() != token.charAt(i))
      return false;
  }

  return true;
}

Capture cam;//Initialise camera in code
//Minim minim;//Initialise minim in code
//AudioPlayer player;//Initialise minim audio player in code
//AudioInput input;

// Global setup
void setup() {
  // Setup graphics
  size(640, 480, OPENGL);
  smooth();
  noStroke();
  frameRate(50);
  //Setup minim audio player
  //minim = new Minim(this);
  //player = minim.loadFile("START.mp3");//load audio file
  //input = minim.getLineIn();

  String[] cameras = Capture.list();

  if (cameras.length == 0)
  {
```

```
    println("There are no cameras available for capture.");
    exit();
  }
  else {
    println("Available cameras:");
    for (int i = 0; i < cameras.length; i++)
    {
      println(cameras[i]);
    }

    // The camera can be initialized directly using an
    // element from the array returned by list():
    cam = new Capture(this, cameras[CAMERA_PORT_NUM]);//selects camera 16 from list
    cam.start();
  }

  output = createWriter(OUTPUT_TXT_FILE_NAME + ".txt");//Name the coords text file according to this final static variable



  // Load font
  font = loadFont("Univers-66.vlw");
  textFont(font);

  // Setup serial port I/O
  println("AVAILABLE SERIAL PORTS:");
  println(Serial.list());
  String portName = Serial.list()[SERIAL_PORT_NUM];
  println();
  println("HAVE A LOOK AT THE LIST ABOVE AND SET THE RIGHT SERIAL PORT NUMBER IN THE CODE!");
  println("  -> Using port " + SERIAL_PORT_NUM + ": " + portName);
  serial = new Serial(this, portName, SERIAL_PORT_BAUD_RATE);
```

```
}

void setupRazor() {
  println("Trying to setup and synch Razor...");

  // On Mac OSX and Linux (Windows too?) the board will do a reset when we connect, which is really bad.
  // See "Automatic (Software) Reset" on http://www.arduino.cc/en/Main/ArduinoBoardProMini
  // So we have to wait until the bootloader is finished and the Razor firmware can receive commands.
  // To prevent this, disconnect/cut/unplug the DTR line going to the board. This also has the advantage,
  // that the angles you receive are stable right from the beginning.
  delay(3000);  // 3 seconds should be enough

  // Set Razor output parameters
  serial.write("#ob");  // Turn on binary output
  serial.write("#o1");  // Turn on continuous streaming output
  serial.write("#oe0"); // Disable error message output

  // Synch with Razor
  serial.clear();  // Clear input buffer up to here
  serial.write("#s00");  // Request synch token
}

float readFloat(Serial s) {
  // Convert from little endian (Razor) to big endian (Java) and interpret as float
  return Float.intBitsToFloat(s.read() + (s.read() << 8) + (s.read() << 16) + (s.read() << 24));
}

void draw() {
  //print coords to .txt
```

```
  output.println("time " + (time/1000));///Convert from ms to seconds
  output.println("yaw "  + (yaw - yawOffset));
  output.println("pitch "  + pitch);
  output.println("roll "  + roll);

   // Reset scene
  background(0);
  lights();

       if (cam.available() == true)
    {
    cam.read();
    }
  image(cam, 0, 0);
    // The following does the same, and is faster when just drawing the image
    // without any additional resizing, transformations, or tint.
    //set(0, 0, cam);
    if (REC_VIDEO == 1)
    {
    saveFrame(OUTPUT_VIDEO_FILE_NAME + frameCount + ".JPEG" );//Record each frame of camera footage with name from this final static variable
    }

  // Sync with Razor
  if (!synched) {
    textAlign(CENTER);
    fill(255);
    text("Connecting...", width/2, height/2, -200);

    if (frameCount == 2)
      setupRazor();  // Set ouput params and request synch token
    else if (frameCount > 2)
      synched = readToken(serial, "#SYNCH00\r\n");  // Look for synch token
```

```
    return;
  }

  // Read angles from serial port USING THESE FOR AHRS RECORDING!!*****************
  while (serial.available() >= 12) {
    time = millis();///<---Timer
    yaw = readFloat(serial);///<---NEW RIGHT HERE!!!!!**
    pitch = readFloat(serial);
    roll = readFloat(serial);
  }

  // Draw board
  pushMatrix();
  translate(width/2, height/2, -350);
  drawBoard();
  popMatrix();

  textFont(font, 20);
  fill(255);
  textAlign(LEFT);

  // Output info text
  text("Point board pins toward screen and press 'a' to align", 10, 25);

  // Output angles
  pushMatrix();
  translate(10, height - 10);
  textAlign(LEFT);
  text("Yaw: " + ((int) (yaw - yawOffset)), 0, 0);
  text("Pitch: " + ((int) pitch), 150, 0);
  text("Roll: " + ((int) roll), 300, 0);
```

```
  popMatrix();
}

void keyPressed()
{
  switch (key)
  {
    case 'z':
      REC_VIDEO = 0;
      output.flush();
      output.close();//push any remaining data to the .txt then close the output writer
    case '0':  // Turn Razor's continuous output stream off
      serial.write("#o0");
      break;
    case '1':  // Turn Razor's continuous output stream on
      serial.write("#o1");
      break;
    case 'f':  // Request one single yaw/pitch/roll frame from Razor (use when continuous
streaming is off)
      serial.write("#f");
      break;
    case 'a':  // Align screen with Razor
      yawOffset = yaw;
//Initialisation indicator addition by CF//
      output.println("*****/////Alignment initialised, video started /////*****");//places a marker
in txt showing alignment start
      REC_VIDEO = 1;
      //player.play();
  }
}
```

Appendix Two: Instructions given to Participants in the Real-World Field Study

Instructions to Candidates

Equipment

You will be asked to wear a backpack containing a computer, and to wear a hat, upon which there will be a number of sensors including a small camera and microphones, and motion detectors.

The motion detector can tell when you are moving, but it cannot track your position or tell anybody where you are.

The camera and microphones will record anything, which is happening near you, and this information will be used only for the purposes of this study and will remain anonymous. It is, however, important that you do not disclose sensitive information whilst wearing the sensor pack (do not use a cash machine, for example).

The researcher will set up the equipment; you do not need to operate it.

Your Task

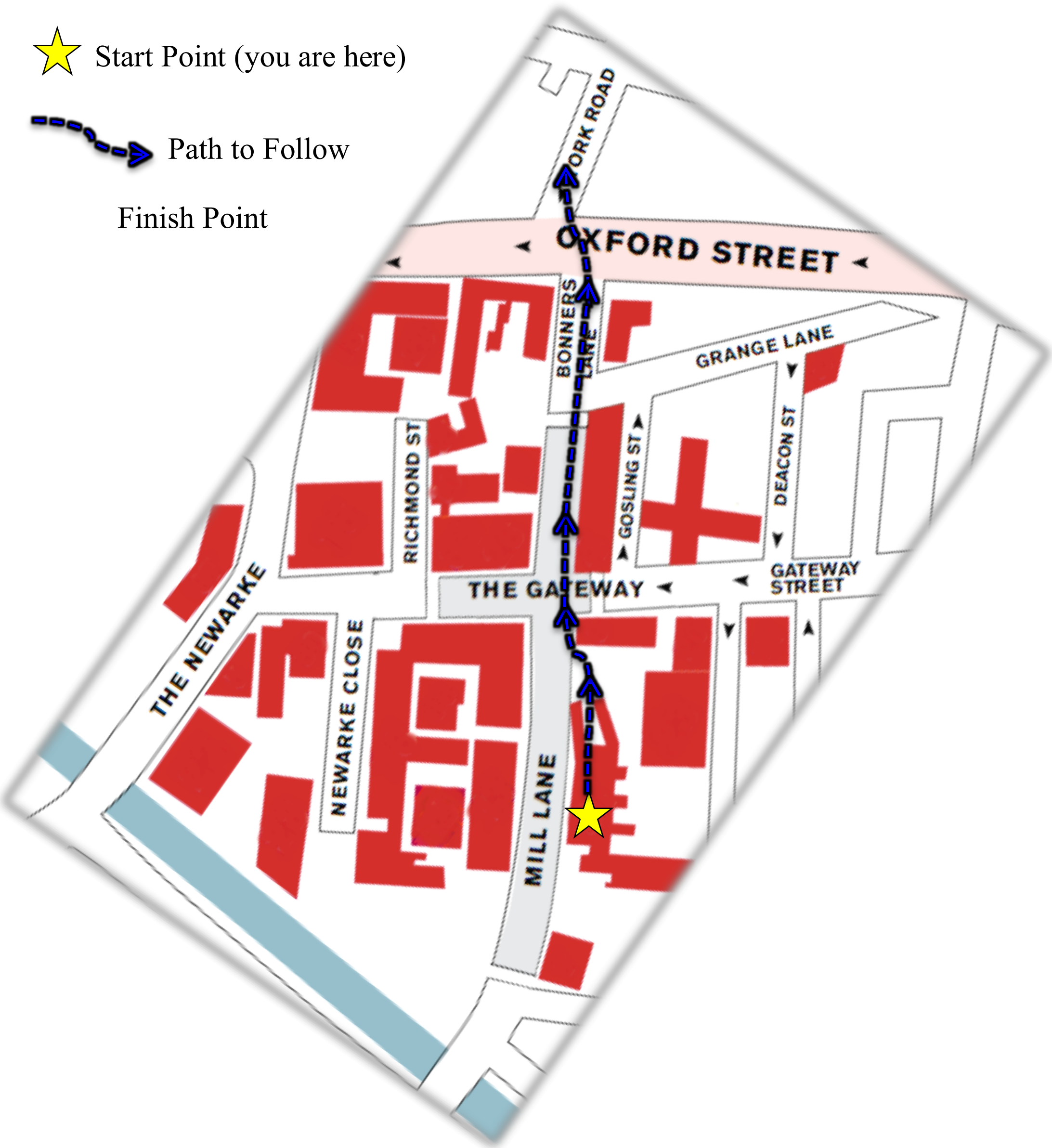
You are asked to follow a short path through the university campus, whilst wearing the sensor package described above. The route will be explained to you, using a map.

During the test we ask that you act as normal, simply walk as you would at any other time, walk along the pavement when possible, and take normal precautions when crossing roads etc.

As you walk the sensor pack will record data about your environment and the sounds around you and save this information to the computer on your back.

If you become uncomfortable during the test, or wish to withdraw from the study, you may remove the equipment and return to the researcher at any time.

NOTE: The researcher will show you how to properly remove the equipment at the beginning of the test.



- Please remain on the pavement whenever possible
- Please be aware that some of the route includes public roads with traffic
- Please use the pedestrian crossing when you reach the main road (Oxford Street) near the end of the route
- A researcher will meet you at the end of the route

Appendix Three: Self-Report Questionnaire (read by researcher) to Participants in the Study

Questionnaire on Navigation Strategies

You may ask for assistance in completing this questionnaire from a friend or helper/assistant if required.

Instructions to assistants:
Please read the questions as they appear on this questionnaire and allow the person you are assisting time to answer for themselves, it is very important that their answers are not altered or influenced by you.

There are instructions on the following pages, which it is important that the person you are assisting understands.

If they are uncomfortable with answering a particular question, or wish to stop completing the questionnaire, then they are free to withdraw or withhold any information as they see fit.

Information for Volunteers
This questionnaire forms part of a research project being undertaken at De Montfort University, Leicester. The aim of the project is to gather information which will be useful to researchers working on assistive technology for people with sensory impairments (particularly visual impairment).

This questionnaire is intended to find out more about how people understand the world around them when navigating outdoors, in particular it is about how we experience sounds and how we understand where they are coming from.

When answering the questions please think about a time, or times, when you were walking somewhere in a busy environment like a street or city/town centre.

The questionnaire starts with some personal questions. These relate directly to the information we are trying to find out, and the answers you provide will be stored securely and anonymously in accordance with the data protection act.

No information which can identify you personally will be stored with this questionnaire, and the answers you give will only be made available to people directly involved in the research project and nobody else.

If you wish to withdraw your participation then all information you have provided us with will be disposed of securely, and no copies will be kept.

If you have any questions, or wish to have the data you have given destroyed please contact:

Chris Feakes (primary researcher) – cfeakes@dmu.ac.uk t:07804598799
Lorenzo Picinali (supervisor) – l.picinali@imperial.ac.uk
Please quote the following participant number in the email: Number ____10____

PEASE KEEP OUR CONTACT INFORMATION AND YOUR PARTICIPANT NUMBER FOR YOUR OWN RECORDS

Statement of Consent
I understand that I may be asked to provide certain pieces of personal information relevant to the research project. This information will be stored anonymously and securely and will be destroyed if I request it.

I understand that my participation is voluntary, and I may withdraw at any time and for any reason, there will be no consequences to me if I choose to withdraw.

I understand that although I will be asked to answer questions based upon my own experiences, I do not need to provide any sensitive information such as where an event took place, or who I was with at the time.

At this stage I am happy to participate in the research project by answering the questionnaire provided. I have the right to withdraw my participation at a later date/time.

Name:________________________

Signature:

____

Date: 07/03/16

Participant Number (to be completed by the researcher): _________

Questions About You

1. Gender (please tick):
Male ☐, Female ☐

2. Age (in years):_______

3. Do you have any visual impairments (please tick)?
Yes ☐, No ☐

4. If yes, which eye is affected (please tick)?
Left ☐, Right ☐, Both ☐

5. Is your visual impairment related to any of the following (please tick)?
Glaucoma ☐, Macular Degeneration ☐, Cataract ☐, Retinopathy ☐, Retinitis ☐, None of these ☐, Don't Know ☐

6. At about what age did your visual impairment begin (in years)?:_____________

7. Could you briefly describe below how your visual impairment affects your eyesight (for example, tunnel vision, clouded vision, no vision etc.):

8. Do you have any hearing impairments (please tick)?
Yes ☐, No ☐

If you have no hearing impairment you may skip to number 14

9. If yes, which ear is affected (please tick)?
Left ☐, Right ☐, Both ☐

10. Is your hearing impairment related to any of the following?
Conductive hearing loss ☐, Sensorineural Hearing Loss ☐, Mixed hearing Loss ☐,
None of these ☐, Don't Know ☐

11. Do you consider yourself to have lost all ability to hear?

Yes □, No □

12. At about what age did your hearing impairment begin (in years)?:_________

Questions About Your Experiences in Busy Environments

13. When answering the following questions please try to think about times when you've had to walk or find your way in environments such streets or town/city centres.

14. In general which of the following do you feel you most rely upon to understand what is happening around you (please tick)?:

Eyesight □, Hearing □, Both □

The following questions are about things that make sounds (for example vehicles, or people)

15. In general do you find it easy to tell in what direction something is by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

16. In general do you find it easy to tell how far away something is by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

17. In general do you find it easy to tell if something is moving towards or away from you by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

18. Is there a particular sound that you find the most difficult to locate in terms of distance or direction? (Please briefly describe it in the space below)

19. Is there a particular sound that you find the easiest to locate in terms of distance or direction? (Please briefly describe it in the space below)

20. Do you use any physical actions or behaviours to help you tell how far away a sound is, for example standing very still, or turning to face the sound (briefly describe them below)?

21. Do you use any physical actions or behaviours to help you tell if a sound is moving towards or away from you (briefly describe them below)?

22. When crossing the road which do you feel is most useful to understand where vehicles are (please tick)?

Eyesight □, Hearing □, Both

23. Do you find it easy to tell in what direction a vehicle is by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know □

24. Do you find it easy to tell how far away a vehicle is by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

25. Do you find it easy to tell if a vehicle is moving towards or away from you by its sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

26. Do you find it easy to tell how many vehicles there are by their sound?

Usually easy □, Sometimes easy □, Often difficult □, Usually difficult □, Don't Know□

27. Do you use any physical actions or behaviours to help you tell how far away a vehicle is by its sound (briefly describe them below)?

28. Do you use any physical actions or behaviours to help you tell if a vehicle is moving towards or away from you by its sound (briefly describe them below)?

29. Do you use any physical actions or behaviours to help you tell how many vehicles there are by their sound (briefly describe them below)?

30. If you have learned any other skills to do with helping you understand the environment you are in by sound, please use the space below to tell us about them:

31. Do you find that the angle of your head affects how well you are able to tell how far away something is by its sound?

Always □, Sometimes □, Not often □, Very Rarely □, Never □

32. Do you ever intentionally move your head to help you tell how far away something is by its sound?

Always □, Sometimes □, Not often □, Very Rarely □, Never □

33. Do you find that the angle of your head affects how well you are able to tell if something is moving towards or away from you by its sound?

Always □, Sometimes □, Not often □, Very Rarely □, Never □

34. Do you ever intentionally move your head to help you tell if something is moving towards or away from you by its sound?

Always □, Sometimes □, Not often □, Very Rarely □, Never □

35. If you do find that moving your head helps you to tell how far away something is, or which direction it is moving by its sound, could you use the space below to describe how you feel it helps?

We appreciate your assistance in completing this questionnaire.

[1] J. Blauert, Spatial Hearing, the Psychophysic of Human Sound Localisation, Cambridge, Massachussetts: The MIT Press, 1996.

[2] B. Katz and L. Picinali, “Spatial Audio Applied to Research with the Blind,” in *Advances in Sound Localisation*, INTECH Open Access, 2011.

[3] E. Wenzel, “Effect of Increasing System Latency on Localization of Virtual Sounds With Short and Long Duration,” in *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, 2001.

[4] S. Werner, G. Gotz and F. Klein, “Influence of Head Tracking on the Externalisation of Auditory Events at Divergence between Synthesised and Listening Room Using Binaural Headphone System,” in *142nd AES Convention*, Berlin, 2017.

[5] G. Wersenyi, “Effect of Emulated Head-Tracking for Reducing Localization errors in Virtual Audio Simulation,” *IEEE Transactions on Audio, Speech and Language Processing,* vol. 17, no. 2, pp. 247-252, 2009.

[6] G. Wersenyi, “Effect of Emulated Head-Tracking for Reducing Localisation Errors in Virtual Audio Simulation,” *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 17, no. 2, pp. 247-252, 2009.

[7] D. Begault, “Auditory and Non-auditory Factors that Potentially Influence Virtual Acoustic Imagery,” in *Audio Engineering Society Conference: 16th International Conference* , 1999.

[8] B. Katz and L. Picinali, “Exploration of Virtual Acoustic Room Simulations by th Visually Impaired,” in *International Seminar on Virtual Acoustics*, Valencia, 2011.

[9] D. Begault, 3-D Sound for Virtual Reality and Multimedia, Moffet Field: National Aeronautics and Space Administration, 2000, pp. 95-96.

[10] RNIB, “Number of Adults and Children Certified with Sight Impairement and Severe SIght Impairement in England and Wales,” 2014. [Online]. Available: http://www.rnib.org.uk/knowledge-and-research-hub/research-reports/general-research/certified-england-wales-12-13. [Accessed 7 April 2017].

[11] B. O'Day, “Employment Barriers for People with Visual Impairements,” *Journal of Visual Impairment and Blindness,* vol. 93, no. 10, 1999.

[12] G. Whitecheck, C. Harrison-Felix, D. Mellick, C. Brooks, B. Charlifue and K. Gerhart, “Quantifying environmental factors: A measure of physical, attitudinal, service, productivity, and policy barriers,” *Archives of Physical Medicine and Rehabilitation,* vol. 85, no. 8, 2004.

[13] J. Swain, S. French, C. Barnes and C. Thomas, Disabling Barriers - Enabling Environments, New York: SAGE, 2013.

[14] R. Parreira, L. Grecco and C. Oliveira, “Postural control in blind individuals: A systematic review,” *Gait and Posture,* vol. 57, pp. 161-167, 2017.

[15] I. Sutherland, “A head-mounted three-dimensional display,” in *In Proceeding of the Fall Joint Computer Conference*, Arlington, VA, 1968.

[16] M. Moshell, “Three views of virtual reality: Virtual environments in the US military,” *Computer,* vol. 26, no. 2, pp. 81-82, 1993.

[17] G. Wersenyi, “Localisation in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications,” in *115th AES Convention* , New York, 2003.

[18] O. Lahav, “Improving orientation and mobility skills through virtual environments for people who are blind; past research and future potential,” in *Proceedings of the 9th International Conference on Disability, Vritual Reality and Associated Technologies*, Lahave, 2012.

[19] L. Picinali, A. Afonso, M. Denis and B. Katz, “Exploration of Architectural Spaces by Blind People Using Auditory Virtual Reality for the Construction of Spatial Knowledge,” *International Journal of Human-Computer Studies,* vol. 72, no. 4, pp. 393-407, 2014.

[20] C. Robinson, S. Mehta and N. Tsingos, “Scalable Format and Tools to Extend the Possibilities of Cinema Audio,” *SMPTE Motion Imaging Journal,* vol. 121, no. 8, 2012.

[21] W. Gardner, “Spatial Audio Reproduction: Towards Individualised Binaural Sound,” in *National Academy of Engineering's Tenth Annual US Frontiers of Engineering Symposium*, 2004.

[22] D. Begault, “Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems,” *Journal of the Audio Engineering Society,* vol. 40, no. 11, pp. 895-903, 1992.

[23] L. Rayleigh, “On our perception of sound direction,” *Philosophy Magazine,* no. 13, pp. 214-232, 1907.

[24] E. Langendijk and A. Bronkhorst, “Contribution of Spectral Cuesto Human Sound Location,” *Journal of the Accoustical Society of America,* pp. 1583-1596, 2002.

[25] P. Bremen, M. Wanrooij and J. v. A. Opstal, “Pina cues Determine Orienting Response Modes to Sychronous Sounds in Elevation,” *The Journal of Neuroscience,* vol. 30, no. 1, pp. 194-204, January 2010.

[26] W. W. Robert A. Lutfi, “Correlational analysis of acoustic cues for the discrimination of auditory motion,” *Journal of the Acoustical Society of America,* pp. 919-928, August 1999.

[27] D. Grantham, B. Hornsby and E. Erpenbeck, “Auditory soatial resolution in horizontal, vertical and diagonal planes,” *Journal of the Acoustical Society of America,* August 2003.

[28] J. C. Middlebrooks and D. M. Green, “Sound localization by human lsteners,” *Annual Reviews Psychology,* pp. 135-159, 1991.

[29] D. Howard and J. Angus, Acoustics and Psychoacoustics, 4th Edition ed., Oxford: Focal Press, 2009.

[30] W. A. Yost and D. M. Green, “Lateralization of Filtered Clicks,” *Journal of the Accoustical Society of America,* no. 50, pp. 1526-1531, 1971.

[31] H. R. Eady and R. G. Klumpp, “Some measurements of interaural time difference thresholds,” *Journal of the Accoustical Society of America,* no. 28, pp. 859-860, 1956.

[32] B. Muller and P. Bovet, “Role of Pinnae and Head Movments in Localising Pure Tones,” *Swiss Journal of Psychology,* vol. 58, no. 3, 1999.

[33] B. Moore, Springer Handbook of Acoustics, T. D. Rossing, Ed., New York: Springer Science + Business Media, 2007.

[34] A. W. Mills, “Lateralization of high frequency tones,” *Journal of the Accoustical Society of America,* no. 32, pp. 132-134, 1960.

[35] J. C. Middlebrooks and D. M. Green, “Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal,” *Journal of the Accoustical Society of America,* no. 86, pp. 89-108, 1989.

[36] A. Pierce, Studies in Auditory and Visual Space Perception, New York: Longmans Green, 1901.

[37] C. Searle, L. Braida, M. Davis and H. Colburn, “Model for Auditory Localisation,” *Journal of the Acoustical Society of America,* vol. 60, no. 1164, 1976.

[38] M. Gardner, “Historical Background of the Haas and/or Precedence Effect,” *Journal of the Accoustical Society of America,* vol. 43, no. 6, pp. 1243-1248, 1968.

[39] W. Yost, “Auditory Image Perception and Analysis: The basis for hearing,” *Hearing Research,* vol. 56, no. 1-2, pp. 8-18, 1991.

[40] A. W. Mills, “On the Minimum Audible Angle,” *The Journal of the Acoustical Society of America,* vol. 30, no. 4, pp. 237-246, April 1958.

[41] D. Perrot and T. J, “Minimum Audible Movement Angle as a Function of Signal Frequency and the Velocity of the Source,” *Journal of the Acoustical Society of America,* pp. 1522-1527, 1988.

[42] D. Perrot, K. Marlborough and P. Merill, “Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate,” *Journal of the Acoustical Society of America,* 1989.

[43] D. R. Perrott and K. Saberi, “Minimum audible angle thresholds for sources varying in both elevation and azimuth,” *Journal of the Accoustical Society of America,* vol. 87, no. 4, pp. 1728-1731, 1990.

[44] W. Grantham, B. Hornsby and E. Erpenbeck, “Auditory Spatial Resolution in Horizontal, Vertical and Diagonal Planes,” *Journal of the Acoustical Society of America,* vol. 114, no. 2, 2003.

[45] A. Lindau and S. Weinzierl, “On the Spatial Resolution of Virtual Acoustic Environments for Head Movments in Horizontal, Vertical, and Lateral Direction,” in *Proceedings of the EAA Symposium on Auralisation*, Espoo, Finland, 2009.

[46] B. Roder, W. Teder-Salejarvl, A. Sterr, F. Rosler, S. A. Hillyard and H. J. Neville, “Improved auditory spatial tuning in blind humans,” *Letters to Nature,* pp. 162-166, 8 July 1999.

[47] J. Garas, Adaptive 3D Sound Systems, Berlin: Springer Science and Business Media, 2012.

[48] R. Easton, A. Greene, P. DiZio and J. Lackner, “Auditory cues for orientation and postural control in sighted and cengenitally blind people,” *Experimental Brain Research,* vol. 118, no. 4, pp. 541-550, 1998.

[49] D. W. Batteau, “The role of the pinna in human localization,” in *Proceedings of the Royal Society of London* , 1967.

[50] H. G. Fisher and S. J. Freedman, “The role of the pinna in auditory localization,” *The Journal of Auditroy Research,* no. 8, pp. 15-26, 1968.

[51] C. L. Searle, L. D. Braida, D. R. Cuddy and M. F. Davis, “Binaural pinna disparity: another localization cue,” *Journal of the Accoustical Society of America,* no. 57, pp. 448-455, 1975.

[52] J. C. Middlebrooks, J. C. Makous and D. M. Green, “Directional sensitivity of sound-pressure levels in the human ear canal,” *Journal of the Accoustical Society of America,* no. 86, pp. 89-106, 1989.

[53] M. Morimoto and K. Nomachi, “On Sound Localisation cues in the Median Plane,” *The Journal of the Acoustical Society of America,* vol. 64, 2005.

[54] D. W. Grantham, B. W. Y. Hornsby and E. A. Erpenbeck, “Auditory spatial resolution in horizontal, vertical, and diagonal planes,” *Journal of the Accoustical Society of America,* vol. 114, no. 2, pp. 1009-1022, 2003.

[55] L. Lida and M. Oota, “Median Plane Sound Localisation using early Head-related Impulse Response,” *Journal of Applied Acoustics,* vol. 139, pp. 14-23, 2018.

[56] D. Perrot, “Concurrent Minimum Audible Angle: A re-examination of the concept of auditory spatial acuity,” *Journal of the Acoustical Society of America,* vol. 75, no. 4, 1984.

[57] V. Best, A. Van Schaik and S. Carlile, “Seperation of concurrent broadband sound sources by human listeners,” *Journal of the Acoustical Society of America,* vol. 115, no. 1, pp. 324-326, 2004.

[58] O. Santala and V. Pulkki, “Directional perception of distributed sound sources,” *Journal of the Acoustical Society of America,* vol. 129, no. 3, pp. 1522-1530, 2011.

[59] T. Kawashima and T. Sato, “Perceptual Limits in Simulated "Cocktail Party",” *Attention, Perception, & Psychophysics,* vol. 77, no. 6, 2015.

[60] X. Zhong and W. Yost, “How Many Images are in an Auditory Scene?,” *Journal of the Acoustical Society of America,* vol. 141, 2017.

[61] S. Handel, Listening: An Introduction to the Perception of Auditory Events, Cambridge, MA: MIT Press, 1989.

[62] N. Durlach and H. Colburn, “Binaural Phenomena,” in *Handbook of Perception*, E. Carterette and M. Friedman, Eds., Academic Press, 1978.

[63] E. Cherry, “Some Experiments on the Recognition of Speech, with one and two Ears,” *The Journal of the Acoustical Society of America,* vol. 25, pp. 975-979, 1953.

[64] G. Kidd, C. Mason, T. Rohtla and P. Deliwala, “release from masking due to spatial seperation of sources in the identification of nonspeech auditory patterns,” *The Journal of the Acoustical Society of America,* vol. 104, no. 422, 1998.

[65] D. Brungart and B. Simpson, “The Effects of Spatial Seperation in Distance on the Informational and Energetic Masking of Nearby Speech Signal,” *The Journal of the Acoustical Society of America,* no. 112, pp. 664-676, 2002.

[66] P. Zahorik, D. Brungart and A. Bronkhorst, “Auditory Distance Perception in Humans: A Summary of Past and Present Research,” *Acta Acustica,* vol. 91, no. 3, pp. 409-420, May/June 2005.

[67] D. H. Mershon and L. E. King, “Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance,” *Perception and Physics,* vol. 18, pp. 409-415, 1975.

[68] G. Parseihian, C. Jouffrais and B. Katz, “Reaching Nearby Sources: Comparison Between Real and Virtual Sound and Visual Targets,” *Frontiers in Neuroscience,* vol. 8, no. 269, 2014.

[69] G. Kearney, M. Gorzel, H. Rice and F. Boland, “distance Perception in Interactive Virtual Acoustic Environents using First and Higher Order Ambisonic Sound Fields,” *Acta Acustica united with Acustica,* vol. 98, pp. 61-71, 2012.

[70] N. Kopco and B. Shinn-Cunningham, “Effect of Stimulus Spectrum on Distance Perception for Nearby Sources,” *The Journal of the Acoustical Society of America,* vol. 130, pp. 1530-1541, 2011.

[71] G. von Békésky, “The Moon Illusion and Similar Auditory Phenomena,” *Amercian Journal of Psychology,* vol. 62, pp. 540-552, 1949.

[72] P. Cochran, J. Thrrop and W. Simpson, “Estimation of Distance of a Source of Sound,” *American Journal of Psychology,* vol. 81, pp. 198-206, 1968.

[73] D. Begault, “Preffered Intensity Increase for Sensation of Half Distance,” *Perceptual and Motor Skills,* vol. 72, pp. 1019-1029, 1991.

[74] D. S. Brungart and K. R. Scott, "The Effects of Production and Presentation Level on Auditory Distance Perception of Speech," *Journal of the Acoustical Society of America,* vol. 110, no. 425, 2001.

[75] D. Mershon and E. King, "Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance," *Perception and Psychophysics,* vol. 18, pp. 409-415, 1975.

[76] J. Petersen, "Estimation of Loudness and Apparent Distance of Pure Tones in a Free Field," *Acustica,* vol. 70, pp. 61-65, 1990.

[77] P. Morse and K. Ingard, Theoretical Acoustics, New York: McGraw-Hill, 1968.

[78] J. Steinberg and B. Snow, "Wire Transmission of Symphonic Music and its Reproduction in Auditory Perspective," New York, 1934.

[79] D. Mershon and J. Bowers, "Absolute and Relative cues for the Auditory Perception of Egocentric Distance," *Perception,* vol. 8, pp. 311-322, 1979.

[80] D. Mershon, W. Ballenger, A. Little, P. McMurtry and J. Buchanan, "Effects of Room Reflectance and Background Noise on Perceived Auditory Distance," *Perception,* vol. 18, pp. 403-416, 1989.

[81] P. Zahorik, "Assessing Auditory Distance Perception using Virtual Acoustics," *The Journal of the Acoustical Society of America,* vol. 111, pp. 1832-1846, 2002.

[82] R. Butler, E. Levy and W. Neff, "Apparent Distance of Sounds Recorded in Echoic and Anechoic Chambers," *Journal of Experimental Psychology,* vol. 6, no. 4, pp. 745-750, 1980.

[83] D. Begault, "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systemd," *Journal of the Audio Engineering Society,* vol. 40, pp. 895-904, 1992.

[84] A. Bronkhorst, "Modeling Auditory Distance Perception in Rooms," AIP Publishing, 2010.

[85] G. v. Bekesky, "The Moon Illusion and Similar Auditroy Phenomena," *The American Journal of Psychology,* vol. 62, no. 4, pp. 540-552, 1949.

[86] A. W. Bronkhorst and T. Houtgast, "Auditory Distance Perception in Rooms," *Nature,* vol. 397, pp. 517-520, 1999.

[87] E. Larsen, N. Iyer, C. Lansing and A. Feng, "On the Minimum Audible Difference in Direct-to-reverberant Energy Ratio," *Journal of the Acoustical Society of America,* vol. 124, pp. 450-461, 2008.

[88] A. Kolarik, B. Moore, P. Zahorik and S. Cristea, "Auditory Distance Perception in Humans: a Review of Cues, Development, Neuronal Bases, and Effects of Sensory Loss," *Attention Perception and Psychophysics,* vol. 78, pp. 373-395, 2015.

[89] P. Zahorik and F. Wightman, "Loudness Constancy with Varying Sound Source Distance," *Nature Neuroscience,* no. 4, pp. 78-83, 2001.

[90] P. Coleman, "Dual Role of Frequency Spectrum in Determination of Auditory Distance," *Journal of the Acoustical Society of America,* vol. 44, 1968.

[91] J. Blauert, Spatial Hearing, Cambridge, MA: MIT Press, 1983.

[92] A. Little, D. Mershon and P. Cox, "Spectral Content as a Cue to Percieved Auditory Distance," *Perception,* vol. 21, pp. 405-416, 1992.

[93] D. Brungart, "Auditory Localisation of Nearby Sources III. Stimulus Effects," *The Journal of the Acoustical Society of America,* vol. 106, pp. 3589-3602, 1999.

[94] D. Brungart, N. Durlach and W. Rabinowitz, "Auditory Localisation of Nearby Sources II. Localisation of a Broadband Source," *The Journal of the Acoustical Society of America,* vol. 106, pp. 1956-1968, 1999.

[95] J. Hass, "An Acoustics Primer," Indiana University, Bloomington, 2003.

[96] R. Holt and W. Thurlow, "Subject Orientation and Judgement of Distance of a Sound Source," *The Journal of the Acoustical Society of America,* vol. 46, August 1969.

[97] D. Greene, "Comment on the Perception of the Range of Sound Source of Unkown Strength," *Journal of the Acoustical Society of America,* vol. 44, p. 634, 1968.

[98] R. Duda and W. Martens, "Range Dependence of the Response of a Spherical HEad Model," *The Journal of the Acoustical Society of America,* vol. 104, pp. 3048-3058, 1998.

[99] B. Shinn-Cunningham and N. Kopco, "Localizing Nearby Sound Sources in a Classroom: Binaural room impulse responses," *Journal of the Acoustical Society of America,* vol. 117, no. 5, 2005.

[100] H. Hirsch, "Perception of the Range of a Sound Source of Unknown Strength," *The Journal of the Acoustical Society of America,* vol. 43, pp. 373-374, 1968.

[101] J. Molino, "Perceiving the Range of a Sound Source when the Direction is Known," *The Journal of the Acoustical Society of America,* vol. 53, pp. 1301-1304, 1973.

[102] E. Wightman and F. Firestone, "Binaural Localisation of Pure Tones," *The Journal of the Acoustical Society of America,* vol. 2, pp. 271-280, 1930.

[103] M. Gardner, "Distance Estimation of 0 degree or Apparent 0 degree Oriented Speech Signals in Anechoic Space," *The Journal of the Acoustical Society of America,* vol. 45, pp. 47-53, 1969.

[104] D. Brungart and W. Rabinowitz, "Auditory Localisation of Nearby Sources. Head Related Transfer Functions," vol. 106, pp. 1465-1479, 1999.

[105] H. Kim, Y. Suzuki, S. Takane and T. Sone, "Control of Auditory Distance Perception based on the Auditory Parallax Model," vol. 62, pp. 245-270, 2001.

[106] P. Zahorik, "Direct-to-reverberant Energy Ratio Sensitivity," *Journal of the Acoustical Society of America,* vol. 112, no. 5, 2002.

[107] M. Lundbeck, G. Grimm, V. Hohmann, S. Luagesen and T. Neher, "Sensitivity to Angular and Radial Source Movmeents as a Function of Acoustic Complexity in Normal and Impaired Hearing," *Trends in Hearing,* vol. 21, pp. 1-14, 2017.

[108] T. Wiesel and H. D.H, "Single-cell Response in Striate Cortex of Kittens Deprived of Vision in one Eye," *Journal of Neurophysiology,* vol. 26, pp. 1003-1017, 1963.

[109] T. Wiesel and D. H. Hubel, "Comparison of the Effects of Unilateral and Bilateral Eye Closure on Cortical Responses in Kittens," *Journal of Neurophysiology,* vol. 28, pp. 1029-1040, 1965.

[110] D. Hubel and T. Wiesel, "The Period of Susceptibility to the Physiological Effects of Unilateral Eye Closure in Kittens," *Journal of Physiology,* vol. 206, no. 2, pp. 419-436, 1970.

[111] D. Bavelier and H. Neville, "Cross-modal Plasticity: Where and How?," *Nature Reviews. Neuroscience ,* vol. 2, pp. 443-452, 2002.

[112] P. Voss, "Sensitive and Critical Periods in Visual Sensory Deprivation," *Frontiers in Psychology,* vol. 4, no. 664, 2013.

[113] J. Rauscheker and M. Korte, "Auditory Compensation of Early Visual Deprivation in the Cat's Anterior Ectosylvian Cortex," *Journal of Neuroscience,* vol. 13, pp. 4538-4558, 1993.

[114] M. Korte and J. Rauschecker, “Auditory Spatial Tuning of Cortical Neuons is Sharpened in Cats With Early Blindness,” *Journal of Neurophysiology,* vol. 70, no. 4, pp. 1717-1721, 1993.

[115] O. Collignon, L. Renier, R. Bruyer, D. Tranduy and C. Veraart, “Improved Selective and Divided Spatial Attnetion in Early Blind Subjects,” *Brain Research,* vol. 1075, pp. 175-182, 2006.

[116] W. Niemeyer and I. Starlinger, “Do the Blind Hear Better? Investigations on Auditory Processing in Congenital or Early Acquired Blindness,” *Audiology,* vol. 20, pp. 510-515, 1981.

[117] G. Fisher, “Spatial Localization by the Blind,” *American Journal of Psychology,* vol. 77, no. 1, pp. 2-14, 1964.

[118] M. Zwiers, A. Van Opstal and J. Cruysberg, “A Spatial Hearing Deficit in Early-blind Humans,” *Journal of Neuroscience,* vol. 21, pp. 1-5, 2001.

[119] N. Lessard, M. Pare, F. Lepore and M. Lassonde, “Early Blind Human Subjects Localize Sound Sources Better than Sighted Subjects,” *Nature,* vol. 395, pp. 278-280, 1998.

[120] B. Roder, A. Teder-Salejarvi, A. Sterr, F. Rosler, S. Hillyard and H. Neville, “Improved Auditory Spatial Tuning in Blind Humans,” *Nature,* vol. 400, pp. 162-166, 1999.

[121] P. Voss, M. Lassonde, F. Gougoux, M. Fortin, J. Guillemot and F. Lepore, “Early and Late-onset Blind Individuals Show Supra-normal Auditory Abilities in far Space,” *Current Biology,* vol. 14, pp. 1734-1738, 2004.

[122] E. Lorenes and L. Renier, “Does Visual Experience Influence the Spatial Distribution of Auditory Attention,” *Acta Psychologica,* vol. 146, pp. 58-62, 2014.

[123] O. Collignon, G. Dormal, G. Allbouy, G. Vandewalle, P. Voss, C. Phillips and F. Lepore, “Impact of Blindness Onset on the Functional Organization and Connectivity of the Occipital Cortex,” *Brain,* vol. 136, no. 9, pp. 2769-2783, 2013.

[124] S. Finocchietti, G. Cappagli and M. Gori, “Encoding Audio Motion: spatial impairement in early blind individuals,” *Frontiers in Psychology,* 2015.

[125] P. Voss, V. Tabry and R. Zatorre, “Trade-Off in the Sound Localisation Abilities of Early Blind Individuals Between the Horizontal and Vertical Planes,” *Journal of Neuroscience,* vol. 35, no. 15, pp. 6051-5056, 2015.

[126] E. Lerens and L. Renier, “Does visual experience influence the spatial distribution of auditory attention?,” *Acta Psychologica,* vol. 146, pp. 58-62, 2014.

[127] E. Lerens and L. Renier, “Does Visual Experience Influence the Spatial Distribution of Auditory Attention?,” *Acta Psychologica,* vol. 146, pp. 58-62, 2014.

[128] P. Voss, M. Lassonde, F. Gougoux, M. Fortin, J. Guillemot and F. Lepore, “Early- and Late- onset Blind Individuals Show Supra-normal Auditory Abilities in far Space,” *Current Biology,* vol. 14, pp. 1734-1738, 2004.

[129] L. Freeman, K. Wood and J. Bizley, “Multisensory Stimuli Improve Relative Localisation Judgements Compared to Unisensory Auditory or Visual Stimuli,” *Journal of the Acoustical Society of America,* vol. 143, 2018.

[130] B. Lok and L. Hodges, “Nerkshire Encyclopedia of Human Computer Interaction,” W. Bainbridge, Ed., Great Barrington MA, Berkshire Publishing LLC, 2004, p. 783.

[131] W. Gaver, “Auditory Icons: Using sound in computer interfaces,” *Human-Computer Interaction,* vol. 2, no. 2, pp. 167-177, 1986.

[132] J. Loomis, R. Golledge and R. Klatzky, “Navigation system for the blind: Auditory display modes and guidance,” *Presence: Teleoperators and Virtual Environments,* vol. 7, no. 2, pp. 193-203, 1998.

[133] F. Zotter and M. Frank, “Ambisonic decoding with panning-invariant loudness on small layouts,” in *144th AES Convention*, Milan, 2018.

[134] B. F. Katz and L. Picinali, “Advances in Sound Localisation,” in *Spatial Audio Applied to Research with the Blind*, INTECH Open Access Publisher, 2011, pp. 225-250.

[135] R. Agarwal and S. Burrus, “Number Theoretic Transforms to Implement Fast Digital Convolution,” *Proceedings of the IEEE,* vol. 63, no. 4, pp. 550-560, 1975.

[136] I. Gel'fand and G. Shilov, Generalized Functions: Properties and Operations, 1st edition ed., Cambridge, MA: Academic Press, 1964.

[137] H. Kuttruff, Room Acoustics, 6th Edition ed., Boca Raton, FL: CRC Press, 2016.

[138] Acoustics Engineering, “Measuring Impulse Responses Using Dirac: An Introduction,” August 2007. [Online]. Available: http://www.acoustics-engineering.com/files/TN001.pdf. [Accessed 26th January 2017].

[139] M. Schroeder, “Integrated-Impulse Method Measuring Sound Decay Without Using Impulses,” *Journal of the Acoustical SOciety of America,* vol. 66, no. 497, 1979.

[140] J. Vanderkooy, “Aspects of MLS Measuring Systems,” *Journal of the Audio Engineering Society,* vol. 42, no. 4, pp. 219-231, 1994.

[141] A. Nobuharu, “Computer-Generated Pulse Signal applied for Sound Measurement,” *The Journal fo the Acoustical Society of America,* vol. 69, no. 5, pp. 1484-1488, 1998.

[142] G. Stan, J. Embrechts and D. Archambeau, “Comparison of Different Impulse Response Measurement Techniques,” *Journal of the Audio Engineering Society,* vol. 50, no. 4, pp. 249-262, 2002.

[143] A. Farina, “Simoultaneous Measurement of Impulse Response and Distortions with Swept-Sine Technique,” Paris, 2000.

[144] AuSIM , “Products. HeadZap,” AuSIM Engineering Solutions, 2011. [Online]. Available: http://ausim3d.com/products/HeadZap/index.html. [Accessed 8 January 2017].

[145] OAW, “Measurement of Head Related Transfer Functions,” Austrian Academy of Sciences, [Online]. Available: https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=article&id=354&Itemid=738&lang=en. [Accessed 4 January 2017].

[146] IRCAM, “Listen HRTF Database,” Istitute for Research and Coordination in Acoustics and Music, 2003. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/. [Accessed 18 December 2016].

[147] CPIC, “The CIPIC HRTF Database,” California University, 2015. [Online]. Available: http://interface.cipic.ucdavis.edu/sound/hrtf.html. [Accessed 3 January 2017].

[148] B. Masiero, Individualised Binaural Technology: Measurment, Equalisation, and Perceptual Evaluation, Berlin: Logos Verlag GmbH, 2012, pp. 47-60.

[149] E. Wenzel, “What Perception Implies about Implementation of Interactive Virtual Acoustic Environments,” in *Audio Engineering Society Conference*, Los Angeles, 1996.

[150] F. Wightman, “Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movmeent,” *The Journal of the Acoustical Society of America,* vol. 105, no. 5, 1999.

[151] E. Wenzel, M. Arruda, D. Kistler and F. Wightman, “Localisation using Non-Individualised Head-Related Transfer Functions,” *Journal of the Acoustical Society of America,* vol. 94, pp. 111-123, 1993.

[152] W. Zhang, M. Zhang, R. Kennedy and T. Abhayapala, “On High-Resolution Head-Realted Transfer Function Measurements: An Efficient Sampling Scheme,” *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 20, no. 2, pp. 575-584, 2011.

[153] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localisation, Cambridge (MA): MIT Press, 1997, pp. 40-55.

[154] H. Langendijk and A. Bronkhorst, “Fidelity of Three-Dimensional-Sound Reproduction using a Virtual Auditory Display,” *The Journal of the Acoustical Society of AMerica,* vol. 107, no. 528, pp. 528-537, 2000.

[155] S. Hosoe, T. Nishino, K. Itou and K. Takeda, “Measurement of the Head-Related Transfer Functions in the Proximal Region,” in *4th European Congress on Acoustics*, Budapest, 2005.

[156] M. Pollow, K. Nguyen, O. Warusfel, T. Carpentier, M. Muller-Trapet, M. Vorlander and M. Noisternig, “Calculation of Head-Related Transfer Functions for Arbitrary Field Points using Spherical Harmonics Decomposition,” *Acta Acustica united with Acustica,* vol. 98, pp. 72-82, 2012.

[157] F. Freeland, L. Biscainho and P. Diniz, “HRTF Interpolation through Direct Angular Parameterisation,” in *IEE International Symposium on Circuits and Systems*, 2007.

[158] D. Kistler and F. Wightman, “A Model of Head-Related Transfer Functions based on Principal Components Analysis and Minimum-Phase Reconstruction,” *The Journal of the Acoustical Society of America,* vol. 91, no. 3, pp. 1637-1647, 1992.

[159] M. Evans, J. Angus and A. Tew, “Analysing Head-Related Transfer Function Measurements using Surface Spherical Harmonics,” *The Journal of the ACoustical SOciety of America,* vol. 104, no. 4, pp. 2400-2411, 1998.

[160] L. Wang, Y. Fuliang and Z. Chen, “Head-Related Transfer Function Interpolation through Multivariate Polynomial Fitting of Principla Component Weights,” *Acoustical Science and Technology,* vol. 30, no. 6, pp. 395-403, 2009.

[161] H. Moller, M. Sorensen, C. Jensen and D. Hammershoi, "Binaural Technique: Do We Need Individual Recordings?," *Journal of the Audio Engineering Society,* vol. 44, pp. 451-469, 1996.

[162] D. Begault, E. Wenzel and M. Anderson, "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualised Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," *Journal of the Audio Engineering Society,* vol. 49, no. 10, pp. 904-916, 2001.

[163] Sony Entertainment, "Playstation VR," 2018. [Online]. Available: https://www.playstation.com/en-gb/explore/playstation-vr/. [Accessed 12 August 2016].

[164] HTC Corporation, "Vive," HTC, 2011. [Online]. Available: https://www.vive.com/uk/. [Accessed 12 August 2016].

[165] Facebook Technologies, "Rift Overview," Occulus, 2016. [Online]. Available: https://www.oculus.com/rift/. [Accessed 12 August 2016].

[166] S. Perrett and W. Noble, "The Contribution of Head Motion Cues to Localization of Low-Pass Noise," *The Journal of Perception and Psychophysics,* vol. 59, no. 7, pp. 1018-26, 1997.

[167] P. Minnaar, O. S.K, F. Christensen and H. Moller, "The Importance of Head Movements for Binaural Room Synthesis," in *Proceedings of the 7th Meeting of the International Conference on Auditory Display*, Espoo, 2001.

[168] G. Wersenyi, "Simulation of Small Head-Movements on a Virtual Audio Display Using Headphone Playback and HRTF Synthesis," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, 2007.

[169] J. Sandvad, "Dynmaic Apsects of Auditory Virtual Reality," in *1ooth Convention of the Audio Engineering Society*, Copenhagen, 1996.

[170] M. Laitinen, T. Pihlajamaki, S. Losler and V. Pulkki, "Influence of Resolution of Head Tracking in Synthesis of Binaural Audio," in *132nd AES Convention*, Budapest, 2012.

[171] H. Wallach, "The Role of Head Movements and Visual Cues in Sound Localization," *Journal of Experimental Psychology,* vol. 27, no. 4, pp. 339-368, 1940.

[172] W. Thurlow and P. Runge, "Effect of Induced Head Movements on Localization of Direction of Sounds," *The Journal of the Acoustical Society of America,* vol. 42, no. 2, pp. 480-488, 1967.

[173] A. Blum, M. Denis and B. Katz, “Navigation in the Absence of Vision: How to find one's way in a 3D audio virtual environment?,” Rome and Perugia, 2006.

[174] D. Perrott, “Auditory and Visual Localisation: two modalities, one world,” in *Audio engineering Society 12th International Conference*, Copenhagen, 1993.

[175] A. Rimell, M. Hollier and R. Voelcker, “The Influence of Cross-Modal Interaction on Audio Visual Speech Quality Perception,” in *Proceedings of the Audio Engineering Society 105th Conference*, San Fransisco, 1998.

[176] M. Hollier and A. Rimmell, “An Experimental Investigation into Multi-Modal Synchronisation Sensitivity for Perceptual Model Development,” in *Proceedings of the Audio Engineering Society 105th Conference*, San Fransisco, 1998.

[177] A. Bronkhorst, “Localisation of Real and Virtual Sound Sources,” *Journal of the Acoustical Society of America,* vol. 98, no. 5, pp. 2542-2553, 1995.

[178] E. Wenzel, “Effect of Increaing System Latency on Localisation of Virtual Sounds,” in *Proceedings of the Audio Engineering Society 16th International Conference on Spatial Sound Reproduction*, Rovaniemi, 1999.

[179] D. Brungart, B. Simpson, R. McKinley, A. Kordik, R. Dallman and D. Ovenshire, “The Interaction Between Head-Tracker Latency, Source Duration, and Response Time in the Localisation of Virtual Sound Sources,” in *Proceedings of ICAD 04- Tenth Meeting* , Sydney, 2004.

[180] Y. Suzuki, S. Yairi and Y. Iwaya, “Effect of Large System Latency of Virtual Auditory Display on Listener's Head Movements in Sound Localisation Task,” in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, 2007.

[181] P. Larsson, D. Vastfjall and M. Kleiner, “Effects of Auditroy Information Consistency and Room Acoustic Cues on Presence in Virtual Environments,” *Acoustical Science and Technology,* vol. 29, no. 2, pp. 191-194, 2008.

[182] R. Clifton, “Breakdown of Echo Suppression in the Precedence Effect,” 1987.

[183] J. Blauert and J. Col, “Irregularities in the Precedence Effect,” in *Auditory Physiology and Perception*, Y. Cazals, K. Horner and L. Demany, Eds., Oxford, Pergamon, 1992, pp. 531-538.

[184] W. Yost and S. Guzman, “Auditory Processing of Sound Sources: Is there an echo in here?,” *Current Directions in Psychological Science,* vol. 5, pp. 125-131, 1996.

[185] D. Begault, 3-D Sound for Virtual Reality and Multimedia, Moffett Field: National Aeronautics and Space Administration, 2000i, pp. 89-94.

[186] M. Schroeder, “Digital Simulation of Sound Transmission in Reverberant Spaces,” *Journal of the Acoustical Society of America,* vol. 47, pp. 424-431, 1970.

[187] W. Hartmann and B. Rackerd, “Localisation of Sound in Rooms II: The Effects of a Single Reflecting Surface,” *Journal of the Acoustical Society of America,* vol. 78, pp. 524-533, 1985.

[188] B. Shinn-Cunningham, “Localising Sound in Rooms,” in Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments, Snowbird, 2001.

[189] M. Cotzin and K. Dallenbach, “"Facial Vision": The Role of Pitch and Loudness in the Perception of Obstacels by the Blind,” *The American Journal of Psychology,* vol. LXIII, no. 4, pp. 485-515, 1950.

[190] W. Kellog, “Sonar System of the Blind,” *Science,* vol. 137, pp. 399-404, 1962.

[191] I. Bassett and E. Eastmond, “Echolocation: Measurement of Pitch versus Distance for Sounds Reflected from a Flat Surface,” *Journal of the Acoustical Society of America,* vol. 36, pp. 911-916, 1964.

[192] C. Rice, “Human Echo Perception,” *Science,* vol. 155, pp. 399-404, 1967.

[193] C. Rice, S. Feinstein and R. Schusterman, “Echo-Detection Ability of the Blind: Size and Distance Factors,” *Journal of Experimental Psychology,* vol. 70, pp. 246-251, 1965.

[194] B. Schenkman and M. Nilsson, “Human Echolocation: Blind and Sighted Persons' Ability to Detect Sounds Recorded in the Presence of a Reflecting Object,” *Perception,* vol. 39, no. 4, pp. 483-501, 2010.

[195] L. Picinali, A. Etherington, C. Feakes and T. Lloyd, “VR Interactive Environments for the Blind: Preliminary Comparative Studies,” in *Joint Virtual Reality Conference 2011 of 8th Conference and Exhibition of the European Association of Virtual Reality and*

Augmented Reality, and the 17th Eurographics Symposium on Virtual Environments, 2011.

[196] D. Rowan, T. Papadopoulos, D. Edwards, H. Holmes, A. Hollingdale, L. Evans and R. Allen, “Identification of the Lateral Position of a Virtual Object based on Echoes by Humans,” *Hearing Research,* vol. 300, pp. 56-65, 2013.

[197] E. Tolman, “Cognitive Maps in Rats and Men,” *The Psychological Review,* vol. 55, no. 4, 1948.

[198] K. Lynch, The Image of the City, Masachusettes: MIT Press, 1960.

[199] O. Lahav and D. Mioduser, “Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind,” *International Journal of Human-Computer Studies,* vol. 66, no. 1, pp. 23-25, 2008.

[200] E. Foulke and P. Hatlen, “A Collaboration of Two Technologies. Part 1: Perceptual and cognitive processes: thier implications for visually impaired persons,” *British Journal of Visual Impairement,* vol. 10, pp. 43-46, 1992.

[201] S. Millar, Understanding and Representing Space: Theory and Evidence from Studies with Blind and Sighted children, Oxford: Oxford University Press, 1994.

[202] C. Thinnius-Blanc and F. Gaunet, “Representation of Space in BLind Persons: vision as a spatial sense?,” *Psychological Bulletin,* vol. 121, pp. 20-42, 1997.

[203] S. von Senden, Space and Sight: the Perception of Space and Shape in the Congenitally Blind Before and After Operation, Glencoe: IL: Free Press, 1932.

[204] R. Kitchin and S. Freundschuh, Cognitive Mapping: Past, Present, and Future, London: Psychology Press, 2000.

[205] O. Depres, V. Candas and A. Dufour, “The estent of visual deficit and auditory spatial compensation: evidence from self-positioning auditory cues,” *Cognitive Brain Research,* vol. 23, no. 2-3, pp. 444-447, 2005.

[206] L. Haber, R. Haber, S. Penningroth, K. Novak and H. Radgowski, “Comparison of 9 methods of indicating the direction to objects- data from blind adults,” *Perception,* vol. 22, pp. 35-47, 1993.

[207] R. Kitchin and R. Jacobson, “Techniques to collect and analyse the cognitive map knowledge of persons with visual impairement or blindness: issues of validity,” *Journal of Visual Impairement and Blindness,* vol. 91, pp. 393-400, 1997.

[208] S. Casey, “Cognitive Mapping by the Blind,” *Journal of Visual Impairement and Blindness,* vol. 72, no. 8, pp. 297-301, 1978.

[209] R. Golledge, R. Klatzky and J. Loomis, “Cognitive mapping and wayfinding by adults without vision,” in *The Construction of Cognitive Maps*, Kluwer Academic Publishers, 1996, pp. 215-246.

[210] S. Ungar, M. Blades and C. Spencer, “The construction of cognitive maps by children with visual impairments,” in *The Construction of Cognitive Maps*, Kluwer Academic Publishers, 1996, pp. 247-273.

[211] K. Papadopoulos, “A school programme contributes to the environmental knowledge of blind people,” *The British Journal of Visual Impairment,* vol. 22, pp. 101-104, 2004.

[212] D. Chebat, S. Maidenbaum and A. Amedi, “Navigation using Sensory Substitution in Real and Virtual Mazes,” *PLOS one,* 2015.

[213] R. Passini, G. Proulx and C. Rainville, “The Spatio-cognitive Abilities of the Visually Impaired Population,” *Environment and Behaviour,* vol. 22, no. 1, 1990.

[214] N. Giudice, R. Klatzky, C. Bennett and J. Loomis, “Perception of 3D location based on vision, touch, and extended touch,” *Experimental Brain Research,* vol. 224, pp. 141-153, 2013.

[215] R. Golledge, R. Klatzky and J. Loomis, “Cognitive Mapping and Wayfinding by Adults without Vision,” in *The Construction of Cognitive MAps*, J. Portugali, Ed., Dordecht/ Boston/ London, Kluwer Academic, 1996, p. 227.

[216] J. Rieser, D. Guth and E. Hill, “Mental processes mediating independent travel: implications for orientation and mobility,” *Journal of Visual Impairment and Blindness,* vol. 76, pp. 213-218, 1982.

[217] J. Rieser, D. Guth and E. Hill, “Sensitivity to perspective structure while walking without vision,” *Perception,* vol. 15, pp. 173-188, 1986.

[218] J. Loomis, R. Klatzky, R. C. J. Golledge, J. Pellegrino and P. Frey, “Nonvisual navigation by blind and sighted- assessment of path integration ability,” *Journal of Experimental Psychology,* vol. 122, pp. 73-91, 1993.

[219] P. Thorndyke and B. Hates-Roth, “Differences in spatial knowledge acquired from maps and navigation,” *Cognitive Psychology,* vol. 14, pp. 560-589, 1982.

[220] B. Witmer, J. Bailey, K. Knerr and K. Parsons, “Virtual Spaces and Real World Places: Transfer of route knowledge,” *Internatioanl Journal of Human-Computer Studies,* vol. 45, no. 4, pp. 413-428, 1996.

[221] R. Darken and J. Sibert, “Wayfinding Strategies and Behaviours in Large Virtual Worlds,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, 1996.

[222] J. Regian, W. Shebilske and J. Monk, “Virtual Reality: An instructional medium for visual-spatial tasks,” *Journal of Communication,* vol. 42, no. 4, 1992.

[223] R. Ruddle, S. Payne and D. Jones, “Navigating buildings in "desk-top" virtual environments: Experimental Invesitgations using extended navigational experience,” *Journal of Experimental Psychology: Applied,* vol. 3, pp. 143-159, 1997.

[224] A. Afonso, B. Katz, A. Blum, C. Jacquemin and M. Denis, “A Study of Spatial Cognition in an Immersive Virtual Audio Environmant: Comparing Blind and Blindfolded Individuals,” in *Proceedings of ICAD 05- Eleventh Meeting of the International Conference on Auditory Display*, Limerick, 2005.

[225] R. Darken and J. Sibert, “Navigating large virtual spaces,” *International Journal of Human-Computer Interaction,* vol. 8, pp. 49-71, 1996.

[226] Seiko Epson, “Gyro Sensors- How the Work and What's Ahead,” Seiko Epson Corp., [Online]. Available: http://www5.epsondevice.com/en/information/technical_info/gyro/. [Accessed 12 December 2016].

[227] M. Steadman, P. Banerjee, S. Nishfuji and S. Tanaka, “Attitude Measurement,” in *The Measurement, Instrumentation, and Sensors: Handbook*, J. Webster, Ed., Springer Science and Business Media.

[228] S. Macintyre, “Magnetic Field Measurement,” in *The Measurement, Instrumentation and Sensors Handbook*, J. Webster, Ed., CRC Press, 1998.

[229] R. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Transactions of the ASME Journal of Basicv Engineering,* vol. 82, pp. 35-45, 1960.

[230] P. Bartz, *9 Degree of Measurement Attitude and Heading Reference System,* Berlin: Deutsche Telekom Laboratories, Quality & Usability Lab, 2011-13.

[231] G. Grossman, R. Leigh, L. Abel, D. Lanska and S. Thurston, "Frequency and Velocity of Rotational Head Perturbations during Locomotion," *Experimental Brain Research,* vol. 70, no. 3, pp. 470-476, 1988.

[232] Mathworks, "Documentation; butter," 2016. [Online]. Available: https://uk.mathworks.com/help/signal/ref/butter.html. [Accessed March 2015].

[233] W. Einhauser, G. Moeller, F. Schuman, J. Conradt, J. Vockeroth, K. Bartl, E. Schneider and P. Konig, "Eye-Head Coordination during Free Exploration in Human and Cat," *Anals of the New York Academy of Sciences,* vol. 1164, no. 1, pp. 353-366, May 2009.

[234] E. Vallabh, Z. Das, M. Zivotofsky, O. Alfred, M. DiScenna and J. Leigh, "Head Perturbations During Walking While Viewing a Head-Fixed Target," *Aviation, Space, and Environmental Medicine,* vol. 66, no. 8, 1995.

[235] H. Burton, "Visual Cortex Activity in Early and Late Blind People," *Journla of NEuroscience,* vol. 23, no. 10, pp. 4005-4011, 2003.

[236] C. Wan, A. Wood, D. Reutens and S. Wilson, "Early but not Late Blindness leads to Enhanced Auditory Perception," *Neuropsychologia,* vol. 48, no. 1, pp. 344-348, 2010.

[237] F. Jiang, G. Stecker, G. Boynton and I. Fine, "Early Blindness Results in Development Plasticity for Auditory Motion Processing within Auditory and Occipital Cortex," *Frontiers in Human Neuroscience,* 2016.

[238] Access Economics Pty Limited, "Future sight loss UK (1): The economic impact of partial sight and blindness in the UK adult population," Royal National Institute for the Blind, 2009.

[239] G. Norman, "Likert scales, levels of measurement and the "laws" of statistics," *Advances in Health Sciences Education,* vol. 15, no. 5, pp. 625-632, 2010.

[240] S.-O. Leung, "A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales," *Journal of Social Service Research,* vol. 37, no. 4, 2011.

[241] W. Albert and T. Tullis, Measuring the User Experience: Collecting, Analysing, and Presenting Usability Metrics, Oxford/Boston: Newnes, 2013.

[242] R. Nisbett and W. T.D, "Telling More Than we can Know: Verbal Reports on Mental Processes," *Psychlogical Review,* vol. 84, no. 3, pp. 231-259, 1977.

[243] C. Carver and M. Scheler, Attention and Self-Regulation: A Control-Theory Approach to Human Behaviour, Berlin: Springer Science and Business Media, 2012.

[244] M. Desmurget and A. Sirigu, “A Parietal-premotor Network for Movement Intention and Motor Awareness,” *Trends in Cognitive Sciences,* vol. 13, no. 10, pp. 411-419, 2009.

[245] A. Tversky and D. Kahneman, “Availability: A heuristic for judging frequency and probability,” *Cognitive Psychology,* vol. 5, no. 2, pp. 207-232, 1973.

[246] H. Kaiser, “An Index of Factorial Simplicity,” *Psychometrika,* vol. 39, no. 1, pp. 31-36, 1974.

[247] B. Efron, “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics,* vol. 7, no. 1, pp. 1-26, 1979.

[248] J. Pezzullo, “The Bootstrap Method for Standard Errors and Confidence Intervals,” [Online]. Available: http://www.dummies.com/education/science/biology/the-bootstrap-method-for-standard-errors-and-confidence-intervals/. [Accessed 24 04 2017].

[249] K. Singh, “On the Asymptotic Accuracy of Efron's Bootstrap,” *The Annals of Statistics,* vol. 9, no. 6, pp. 1187-1195, 1981.

[250] P. Bickel and D. Freedman, “Some Asymptotic Theory for Bootstrap,” *The Annals of Statistics,* vol. 9, pp. 1196-1217, 1981.

[251] T. Hesterberg, “What Teachers should know about the Bootstrap: resampling in the undergraduate statistics curriculum,” *The American Statistician,* vol. 69, no. 4, pp. 371-386, 2015.

[252] B. Brightwell and B. Dransfield, Avoiding and Detecting Statistical Malpractice, eBook: InfluentialPoints, 2013.

[253] B. Efron, “The Jackknife, the Bootstrap, and Other Resampling Plans,” SIAM, 1982.

[254] D. Wright, K. London and A. Field, “Using Bootstrap Estimation and the Plug-in Principle for Clinical Psychology Data,” *Journal of Experimental Psychopathology,* vol. 2, no. 2, pp. 252-270, 2011.

[255] D. Withington, “Localisable Alarms,” in *Human Factors in Auditory Warnings*, Ashgate, 1999, pp. 33-40.

[256] L. Picinali, “Techniques for the Extraction of the Impulse Response of a Liner and Time-invariant System,” in *Proceedings of the DMRN Doctoral Research Conference* , London, 2006.

[257] GRAS Sound and Vibration, “KEMAR Manikin Type 45BA Product Data and Specifications,” GRAS Sound and Vibration, 12 October 2007. [Online]. Available: http://www.ee.bgu.ac.il/~acl/Equip/KEMAR.pdf. [Accessed 17 June 2016].

[258] T. Scgultz, “Relationship between Sound Power Level and Sound Pressure Level in Dwellings and Offices,” *ASHRAE Transactions,* vol. 91, no. 1, 1985.

[259] V. Raykar, R. Duraiswami and B. Yegnanarayana, “Extracting Frequencies of the Pinna Spectral Notches in Measured Head Related Impulse Responses,” *Journal of the Acoustical Society of America,* vol. 118, no. 1, pp. 364-374, 2005.

[260] R. Butler and K. Belendiuk, “Spectral Cues Utilized in the Localization of Sound in the Median Sagital Plane,” *Journal of the Acoustical Society of America,* vol. 61, 1977.

[261] S. Klockgether and S. van de Par, “Just Noticable Differences of Spatial Cues in Echoic and Anechoic Acoustical Environments,” *Journal of the Acoustical Society of America,* vol. 140, no. 4, 2016.

[262] M. Mascia, A. Canclini, F. Antonacci, M. Tagliasacchi, A. Sarti and S. Tubaro, “Forensic and anti-forensic analysis of indoor/outdoor classifiers based on acoustic clues,” in *23rd European Signal Processing Conference*, Nice, 2015.

[263] A. Brughera, “Human Interaural Time Difference Thresholds for Sine Tones: The high frequency limit,” *Journal of the Acoustical Society of America,* vol. 133, no. 5, pp. 2839-2855, 2013.

[264] A. Lindau, “The Perception of System Latency in Dynamic Binaural Synthesis,” in *NAG/DAGA 2009*, Rotterdam, 2009.

[265] Sparkfun Electronics US, “9 Degrees of Freedom Razor IMU,” [Online]. Available: https://www.sparkfun.com/products/retired/10736. [Accessed Febuary 2015].

[266] Wikipedia Foundation, “Butterworth Filter,” 2016. [Online]. Available: https://en.wikipedia.org/wiki/Butterworth_filter. [Accessed 12 January 2017].

[267] K. Riederer, “Transfer Function Measurements in Audio,” Espoo, 1996.

[268] The Audio Engineering Society, "AES Recommended Practice for Professional Digital Audio- Preferred Sampling Frequencies for Applications Employing Pulse-Code Modulation," 2014.

[269] E. Newman, "Speech and Hearing," in *American Institute of Physics Handbook*, New York, McGraw-Hill, 1972, pp. 3-155.

[270] N. Macmillan and C. Creelman, Detection Theory: a User's Guide, Mahwah: Lawrence Erlbaum Associates, Inc., 2005.

[271] R. Lowry, "Concepts and Applications of Inferential Statistics," 2017. [Online]. Available: http://vassarstats.net/textbook/index.html. [Accessed 4 March 2017].

[272] P. Majdac, B. Laback, M. Goupell and M. Mihocic, "The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Virtual Environment," in *Audio Engineering Society 124th Convention*, Amsterdam, 2008.

[273] A. Musicant and R. Butler, "The influence of pinnae-based spectral cues on sound localization," *Journal of the Acoustical Society of America,* vol. 75, no. 1195, 1984.

[274] D. Perrot, "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial accuity," *Journal of the Accoustical Society of America,* vol. 75, no. 4, pp. 1201-1206, 1984.

[275] L. Picinali, "The creation of a binaural spatialization tool," 2011.